TEACHING INTRODUCTORY PSYCHOLOGY:
TIPS FROM TOP

Edited by
Richard A. Griggs
University of Florida

Sherri L. Jackson
Jacksonville University
FOREWORD

This book follows in the footsteps of the first three volumes in the Handbook for Teaching Introductory Psychology series. In the prefaces to these volumes, the various editors all stressed two major points relevant to the development of this series. These comments also apply to this book. First, introductory psychology is one of the most popular courses with students, and it is central in importance to the undergraduate psychology curriculum. Second, Teaching of Psychology (ToP) is recognized as one of the premier disciplinary pedagogical journals and, as such, regularly includes excellent articles on teaching the introductory course. Thus, a current, readily accessible, and organized collection of articles from ToP on teaching the introductory course should be an invaluable resource to introductory teachers, and indeed, this has proved to be the case for the first three volumes. We hope that the same holds true for this volume. To emphasize the resource nature of this new volume in the series, we abandoned the more formal previous titles and entitled this new volume, Teaching Introductory Psychology: Tips from ToP.

The organizational structure of the first three volumes was relatively the same. The editors presented the articles in two major sections (Issues and Approaches in Teaching Introductory Psychology and Demonstrations and Activities in Introductory Psychology). The subsection topics in the first section have remained fairly constant across volumes, with some new topics on assessment added to Volume III. These topics include not only general approaches to teaching the introductory course but also more specific topics such as introductory textbook selection, all aspects of testing ranging from student preparation to testing factors, and extra credit. The ordering of the subsection topics in the Demonstrations and Activities section parallels the typical chapter-topic sequence used in most introductory psychology textbooks. To maintain continuity across volumes, we incorporated this organizational structure for these two sections in the present volume. Most of the subsections from the first three volumes appear in this one. However, because the number of articles on the role of technology in teaching the introductory course (e.g., online versions of the course, the use of computerized quizzing, and use of electronic student response systems) has dramatically increased over the past decade due to the importance of the role of technology in teaching, we include a new major section (Technology in Teaching Introductory Psychology) devoted to them. This new section has been positioned in between the other two major sections.

With the first issue of ToP appearing in 1974 and the third volume in this handbook series including articles up to 2000, the average time period covered for each volume is a little
over eight years. This volume, however, covers a slightly longer publication period, from 2000 to 2010. We conducted a thorough examination of all of the issues of ToP during this period and identified over 170 articles relevant to teaching introductory psychology to evaluate further for possible inclusion in this volume. Of these, we chose 113 articles for inclusion—61 for Issues and Approaches in Teaching Introductory Psychology, 25 for Technology in Teaching Introductory Psychology, and 27 for Demonstrations and Activities in Introductory Psychology. The articles within each subsection are ordered chronologically according to their publication dates starting with the earliest one. The citation information for each article is provided in the Citation Index.

Each of the three major sections is broken down into more specific subsections so that you can go directly to the topic of your choice. For example, the new Technology in Teaching Introductory Psychology section has been divided into six subsections: Online Delivery of the Course, Computer Assisted Instruction, Posting Course Lecture Notes, Electronic Review Tools, Electronic Student Response Systems, and Computerized Quizzing. Thus, for example, if you are interested in using Clickers in your classroom, you can go to the Electronic Student Response Systems subsection to see what the ToP authors on this topic have learned from their use of and research on such systems. We also provide a Brief Table of Contents so that you can more easily view the entire organizational structure of the book—the three major sections with all of the subsections for each one listed. In the Demonstrations and Activities section, you will find that we have not provided a demonstration or activity for each major chapter topic in an introductory textbook because during the period covered by this volume, no relevant demonstrations or activities for some topics were published in ToP. This section, however, still provides 27 excellent demonstrations or activities for 10 different introductory course topics.

As with the previous three volumes, this book contains the work of many dedicated teachers. We would like to thank all of them for sharing their ideas and research with us. We would also like to thank Bill Buskist and Jeff Stowell for their invaluable help in bringing this project to fruition.

Richard A. Griggs
University of Florida

Sherri L. Jackson
Jacksonville University
# BRIEF TABLE OF CONTENTS

## Section I. Approaches and Issues in Teaching Introductory Psychology
1. Approaches to the Introductory Course
2. Research Participation for Introductory Students
3. Active Learning
4. Examinations: Test Items
5. Examinations: Student Preparation
6. Examinations: Testing Factors
7. Examinations: Course Design Factors
8. Examinations: Effects of Study Guide and Pedagogical Aid Use
9. Students’ Course Preferences and Knowledge about Psychology
10. Introductory Textbooks: Selection and Student Use
11. Introductory Textbooks: Content
12. Extra Credit

## Section II. Technology in Teaching Introductory Psychology
1. Online Delivery of the Course
2. Computer Assisted Instruction
3. Posting Course Lecture Notes
4. Electronic Review Tools
5. Electronic Student Response Systems
6. Computerized Quizzing

## Section III. Demonstrations and Activities in Introductory Psychology
1. General
2. Research Methods and Statistics
3. Sensation and Perception
4. Learning
5. Memory
6. Cognition
7. Social Psychology
8. Personality
9. Abnormal Psychology
10. Industrial-Organizational Psychology

## Appendix: Citation Information
TABLE OF CONTENTS

Front Matter and Appendix

Title Page
Foreward
Brief Table of Contents
Full Table of Contents
Appendix: Citation Information

Section I. Approaches and Issues in Teaching Introductory Psychology

1. Approaches to the Introductory Course

Service learning in a general psychology class: Description, preliminary evaluation, and recommendations. Molly D. Kretchmar

Instructional television versus traditional teaching of an introductory psychology course. Steven F. Bacon and Julie A. Jakovich

Exposure to the fields of psychology: Evaluation of an introductory psychology project. Amanda M. Maynard, Douglas C. Maynard, and Kirsten A. Rowe

Introductory psychology topics and student performance: Where’s the challenge? Andrew C. Peck, Rahan S. Ali, Robert L. Matchock, and Max E. Levine

Techniques for increasing student learning from educational videos: Notes versus guiding questions. Timothy J. Lawson, James H. Bodle, and Tracy A. McDonough

Classic articles as primary source reading in introductory psychology. Richard A. Griggs and Sherri L. Jackson

The effect of refuting misconceptions in the introductory psychology class. Patricia Kowalski and Annette Kujawski Taylor
Do student perceptions of diversity emphasis relate to perceived learning of psychology. Joelle D. Elicker, Andrea F. Snell, and Alison L. O’Malley

Does the first week of class matter? A quasi-experimental investigation of student satisfaction. Anthony D. Hermann, David A. Foster, & Erin E. Hardin

Exploring interdisciplinary themes in introductory psychology. Kristin A. Ritchey & Jennifer P. Bott

2. Research Participation for Introductory Students

Using a dining facility as an introductory psychology research laboratory. Nancy Koschmann and Richard Wesp

Using exam bonus points as incentive for research participation. Joseph R. Ferrari and Stephanie McGowan

An argument for a laboratory in introductory psychology. Howard C. Berthold, Christopher M. Hakala, and Dennis Goff

The pedagogical value of experimental participation paired with course content. Michelle Ceynar Rosell, Danielle M. Beck, Katie E. Luther, Kelly M. Goedert, Wendelyn J. Shore, and Dana D. Anderson

Introductory psychology students’ perceptions of alternatives to research participation. David Trafimow, Laura Madson, and Iola Gwizdowski

Introducing students to psychological research: General psychology as a laboratory course. Thomas J. Thieman, E. Gil Clary, Andrea M. Olson, Rachel C. Dauner, and Erin Ring

A video introduction to psychology: Enhancing research interest and participation. Donald F. Sacco and Michael J. Bernstein
3. **Active Learning**

Active learning within a lecture: Assessing the impact of short, in-class writing exercises. *Adam Butler, Kayah-Bah Phillmann, and Lona Smart*

Obedience, conformity, and social roles: Active learning in a large introductory psychology class. *April L. Bleske-Rechek*

Focused interactive learning: A tool for active class discussion. *Helen C. Harton, Deborah S. Richardson, Ricardo E. Barreras, Matthew J. Rockloff, and Bibb Latané*

Using case studies in introductory psychology. *Julie A. Leonard, Kirsten L. Mitchell, Steven A. Meyers, and Jacqueline D. Love*

A motivating exercise for the introductory class (and beyond). *Louise Katz*

Students teaching students: An experiential learning opportunity for large introductory psychology classes in collaboration with local elementary schools. *Gary M. Muir and Gretchen J. van der Linden*

4. **Examinations: Test Items**

Difficulty and discriminability of introductory psychology test items. *Charles Scialfa, Connie Legare, Larry Wenger, and Louis Dingley*

Using ignorance questions to promote critical thinking skills. *David W. Carroll*

The use of discrimination indexes in constructing course exams: A question of assumptions. *Daniel R. Stalder*

5. **Examinations: Student Preparation**

Study tips: How helpful do introductory psychology students find them? *William R. Balch*

Student perspectives on grade changes from test to test. *Baron Perlman and Lee I. McCann*
Improving students’ exam performance by introducing study strategies and goal setting. Victoria Manion Fleming

How do students really study (and does it matter)? Regan A. R. Gurung

Academic background and course involvement as predictors of exam performance. Byron L. Zamboanga, Laura M. Padilla-Walker, Sam A. Hardy, Ross A. Thompson, and Sherry C. Wang

Effects of test expectation on multiple-choice performance and subjective ratings. William R. Balch

6. Examinations: Testing Factors

Students’ reasons for writing on multiple-choice examinations. Frank M. LoSchiavo and Mark A. Shatz


Effect of crib card construction and use on exam performance. K. Laurie Dickson and Michelle D. Miller


7. Examinations: Course Design Factors

The exam-a-day procedure improves performance in psychology classes. Frank C. Leeming

Influence of unannounced quizzes and cumulative exam on attendance and study behavior. Haig Kouyoundjian

Elaborations of introductory psychology terms: Effects on test performance and subjective ratings. William R. Balch

The impact of daily extra credit quizzes on exam performance. Laura M. Padilla-Walker
Introductory psychology student performance: Weekly quizzes followed by a cumulative final exam. *R. Eric Landrum*

8. **Examinations: Effects of Study Guide and Pedagogical Aid Use**

Pedagogical aids and student performance. *Regan A. R. Gurung*

Pedagogical aids: Learning enhancers or dangerous detours? *Regan A. R. Gurung*

Effect of textbook study guides on student performance in introductory psychology. *K. Laurie Dickson, Michelle D. Miller, and Michael S. Devoley*

Effect of study guide exercises on multiple-choice exam performance in introductory psychology. *K. Laurie Dickson, Michael S. Devoley, and Michelle D. Miller*

9. **Students’ Course Preferences and Knowledge about Psychology**

Grade expectations. *Jane F. Gaultney and Arnie Cann*

Prior knowledge and its relevance to student achievement in introduction to psychology. *Ross A. Thompson and Byron L. Zamboanga*

Does deliberate source monitoring reduce students’ misconceptions about psychology? *Joshua D. Landau and Anthony J. Bavaria*

Empowering students: Class-generated course rules. *Jeannie D. DiClementi and Mitchell M. Handelsman*

10. **Introductory Textbooks: Selection and Student Use**

Textbook selection: Balance between the pedagogy, the publisher, and the student. *R. Eric Landrum and LuAnne Hormel*

Student use of introductory texts: Comparative survey findings from two universities. *Jason F. Sikorski, Kelly Rich, Bryan K. Saville, William Buskist, Oksana Drogan, and Stephen F. Davis*
Using a core textbook for the introductory course. Richard A. Griggs, Sherri L. Jackson, and Pam Marek

Using common core vocabulary in text selection and teaching the introductory course. Richard A. Griggs, Alexandra Bujak-Johnson, and Derrick L. Proctor

Evaluating the electronic textbook: Is it time to dispense with the paper text? James A. Shepperd, Jodi L. Grace, and Erika J. Koch

11. Introductory Textbooks: Content

Similarity of introductory psychology textbooks: Reality or illusion. Richard A. Griggs and Pam Marek

Operant conditioning concepts in introductory psychology textbooks and their companion web sites. Jane P. Sheldon

Psychology textbooks: Examining their accuracy. Faye B. Steuer and K. Whitfield Ham, II

The representation of applied psychology areas in introductory psychology textbooks. Charlotte W. Haselhuhn and Kerri L. Clopton

12. Extra Credit

Extra credit exercise: A painless pop quiz. B. Michael Thorne

Extra credit: Gifts for the gifted? Marjorie S. Hardy

Breaking the silence: Using a token economy to reinforce classroom participation. Kurt A. Boniecki and Stacy Moore

Section II. Technology in Teaching Introductory Psychology

1. Online Delivery of the Course

The online delivery of psychology courses: Attrition, performance, and evaluation. Stefanie B. Waschull
Integrating technology and pedagogy: Web instruction and seven principles of undergraduate education. *Michael H. Newlin and Alvin Y. Wang*

Teaching in cyberspace: Online versus traditional instruction using a waiting-list experimental design. *Christopher R. Poirier and Robert S. Feldman*

Predicting success in online psychology courses: Self-discipline and motivation. *Stefanie B. Waschull*

Enhancing online instruction with humor. *Frank M. LoSchiavo and Mark A. Shatz*

2. **Computer Assisted Instruction**

Are computer-assisted teaching methods effective? *Kurt A. DeBord, Mara S. Aruguete, and Jeannette Muhlig*

Employing computer-administered exams in general psychology: Student anxiety and expectations. *Carolyn A. Schult and John L. McIntosh*

Using group web page and video clip creation exercises in introductory psychology courses. *Terry F. Pettijohn II and Elizabeth G. Perelli*

Presentation software in the college classroom: Don’t forget the instructor. *Erin E. Hardin*

Does an interactive WebCT site help students learn? *Joelle D. Elicker, Alison L. O’Malley, and Christine M. Williams*

They hear, but do not listen: Retention for podcasted material in a classroom context. *David B. Daniel & William Douglas Woody*

3. **Posting Course Lecture Notes**

Providing students with instructors’ notes: Problems with reading, studying, and attendance. *Michael A. Vandehey, Crystale M. Marsh, and George M. Diekhoff*

Differential effects of full and partial notes on learning outcomes and attendance. *Tara L. Cornelius & Jamie Owen-DeSchryver*
If you post it, will they come? Lecture availability in introductory psychology. *M. Christina Hove and Kevin J. Corcoran*

4. **Electronic Review Tools**

Using interactive computer technology to enhance learning. *Joy R. Pemberton, Joaquin Borrego, Jr., and Lee M. Cohen*

A technology classroom review tool for general psychology. *Stephen T. Paul, John A. Messina, and Alma M. Hollis*

5. **Electronic Student Response Systems**

Promoting active learning using individual response technology in large introductory psychology classes. *Christopher R. Poirer and Robert S. Feldman*

Benefits of electronic audience response systems on student participation, learning, and emotion. *Jeffrey R. Stowell and Jason M. Nelson*

Using wireless response systems to replicate behavioral research findings in the classroom. *Anne M. Cleary*

Efficacy of personal response systems (“clickers”) in large, introductory psychology classes. *Beth Morling, Meghan McAuliffe, Lawrence Cohen, and Thomas M. DiLorenzo*

Using student response systems (“Clickers”) to combat conformity and shyness. *Jeffrey R. Stowell, Terrah Oldham, & Dan Bennett*

6. **Computerized Quizzing**

Effective student use of computerized quizzes. *Thomas Brothen and Cathrine Wambach*

The value of time limits on internet quizzes. *Thomas Brothen and Cathrine Wambach*

Are online study questions beneficial? *Kristin Grimstad and Mark Grabe*
Using web-based quizzing to improve exam performance: Lessons learned.
*David B. Daniel and John Broida*

**Section III Demonstrations and Activities in Introductory Psychology**

1. **General**

   Forbidden words: A strategy for studying psychology. *Michelle M. Merwin*

   Encouraging distributed study: A classroom experiment on the spacing effect. *William R. Balch*

   Improving students’ study habits by demonstrating the mnemonic benefits of semantic processing. *Julie M. Bugg, Edward L. DeLosh, and Mark A. McDaniel*

2. **Research Methods and Statistics**

   A one-minute “intelligence” test. *Richard A. Griggs*

   A psychic-reading demonstration designed to encourage critical thinking. *Timothy J. Lawson*

   Teaching the principles of test validation in introductory psychology. *Richard Wesp and Sussie Eshun*

   Introducing psychology students to research methodology: A word-pleasantness experiment. *William R. Balch*

   Demonstrating experimenter “ineptitude” as a means of teaching internal and external validity. *Kimberli R. H. Treadwell*

3. **Sensation and Perception**

   Classroom demonstrations of auditory perception. *LaDawn Haws and Brian J. Oppy*

4. **Learning**


5. **Memory**

   Examining memory phenomena through flashbulb memories. *Mark Sudlow Hoyert and Cynthia D. O’Dell*
   
   An active learning classroom activity for the “cocktail party phenomenon.” *Michael A. Clump*

6. **Cognition**

   An effective exercise for teaching cognitive heuristics. *Alan Swinkels*
   
   Helping students gain insight into mental set. *Richard A. Griggs*
   
   Demonstrating the Monty Hall dilemma. *Matthew R. Kelley*

7. **Social Psychology**

   Demonstrating the concept of illusory correlation. *Jay W. Jackson*
   
   “Me conform? No way”: Classroom demonstrations for sensitizing students to their conformity. *C. R. Snyder*
   
   Using a “new classic” film to teach about stereotyping and prejudice. *Andrew N. Christopher, Jamie L. Walter, Pam Marek, and Cynthia S. Koenig*
   
   You are what you wear: An interactive demonstration of the self-fulfilling prophecy. *Michelle R. Hebl and Eden B. King*
   
   Using *The Simpsons* to teach social psychology. *Judy Eaton and Ayse K. Uskul*
8. **Personality**

What’s in a name? Better letters if it’s mine! *Angela Lipsitz and Lance A. Gifford*

Why does the “above average effect” exist? Demonstrating idiosyncratic trait definition. *Jason A. Nier*

I scream, you scream: Teaching validity and reliability via the ice cream personality test. *Marianne Miserandino*

Heeeere’s Johnny: A case study in the five factor model of personality. *Marianne Miserandino*

9. **Abnormal Psychology**

Teaching students to evaluate web information as they learn about psychological disorders. *Mark A. Casteel*

10. **Industrial/Organizational Psychology**

An evaluation of industrial/organizational psychology teaching modules for use in introductory psychology. *Douglas C. Maynard, Peter D. Bachiochi, and Ana C. Luna*
Service Learning in a General Psychology Class: Description, Preliminary Evaluation, and Recommendations

Molly D. Kretchmar
Gonzaga University

In this article, I describe a service-learning project for an introductory psychology course. Thirty-two of 36 general psychology students chose service learning over an optional research project, and most students (over 80%) evaluated the experience favorably in terms of its impact on their learning and commitment to service. The discussion focuses on recommendations about issues ranging from the logistics of developing and managing a service-learning course to concerns inherent in placing students in community agencies.

When outstanding teachers are interviewed or students are asked about their most meaningful learning experiences, they often highlight the active engagement and participation of the student in the learning process (see Halonen, 1992; 1997 Teaching Award Winners, 1997). Additionally, recent years have witnessed a growing concern about the education of students as ethically responsible citizens. In 1985, the presidents of Brown, Georgetown, and Stanford Universities along with the Education Commission of the States founded Campus Compact, an organization dedicated to fostering among students the development of “values and skills of civic participation through involvement in public service” (Kobrin & Mareth, 1996, p. 8). More specific to the discipline of psychology, Brewer et al. (1993) articulated 11 recommendations for the psychology curriculum, including “ensure that courses more accurately reflect the diversity of humankind”; “provide students with the experience and understanding they will need to make the world a better place in which to lead productive and fulfilling lives”; and “help students understand psychology as a science, a profession, and a means of promoting human welfare” (pp. 179–180). Service learning, which integrates community-based, hands-on service into an organized course, is one pedagogical approach that encompasses a number of these goals and has the potential to transform both students’ intellect and their character.

Fundamentally different from volunteer work, service learning emphasizes reciprocity by creating a learning opportunity for students while also serving the needs of a community group or agency. Both elements—service and learning—are highlighted: “The service reinforces and strengthens the learning, and the learning reinforces and strengthens the service” (Cooper, 1996a, p. 1). Moreover, community service often follows a charity-based model in which there can be a patronizing distinction between those serving and those seeking services. In contrast, service learning emphasizes an equal exchange: “Both the server and those served teach, and both learn” (Kendall, 1991, p. 20). Finally, service learning goes beyond traditional community service by connecting the volunteer experience to specific academic goals, which are facilitated by reflection, discussion, and integration with course material (Cooper, 1996b; Kendall, 1991; Kobrin & Mareth, 1996; Morton, 1993; see also Dunlap, 1998).

Service learning is positively associated with a variety of student outcomes ranging from intellectual growth to social action (Batchelder & Root, 1994; Giles & Eyler 1994; Gray et al., 1996; Markus, Howard, & King, 1993). For example, a large-scale evaluation of the initiative Learn and Serve America, Higher Education (Gray et al., 1996) reported higher levels of academic achievement (measured by multiple indexes including grades, degree aspirations, gains in knowledge, time devoted to academic studies, and completion of extra work) among service participants than among nonparticipants. Service participants also scored higher than their nonparticipant counterparts on a variety of measures of civic responsibility (e.g., commitment to helping others in difficulty) and on indexes of life skills (e.g., understanding community problems). These group differences remained significant even after controlling for student characteristics assessed prior to service learning, such as the predisposition to engage in service. Gray et al.’s findings were further substantiated by Markus et al. (1993), who found that students randomly assigned to “community service” sections in an introductory politics class earned a higher final course grade and scored higher on a self-report of learning and on a number of markers of personal change (e.g., awareness of social problems and personal responsibility) than did students in “traditional” sections (pp. 412–413).

The value of service learning in enhancing academic skills, in increasing awareness of the diversity of humanity and the complexity of human hardship, and in fostering one’s civic responsibility is both theoretically and empirically justified. In this article, I provide a description of a service-learning project that I implemented in an introductory psychology course as well as a descriptive evaluation of the project. The discussion focuses on recommendations for service learning based on the students’ evaluations and my experiences.
The Service-Learning Project

During the Fall 1997 semester students enrolled in my section of General Psychology completed either a service-learning project or a research project as part of the course requirements. At the beginning of the semester I explained the requirements of each project and provided students interested in service learning with a complete written description of the project, along with a list of possible agencies with whom the university’s service-learning coordinator had already established relationships. Students secured their own placements and, with input from their agency supervisors, completed their service-learning contracts, which specified hours scheduled, learning objectives, and activities.

Given the introductory nature of the course and students’ time constraints, I required that students complete a minimum of 15 hr across the semester. Their schedules, negotiated with agency supervisors, were flexible (e.g., 1 to 2 hr once a week, 3 to 4 hr twice a month) as long as both student and agency needs would be met. To integrate their service-learning activities with class material, students drew on their experiences during small and large group discussions around questions designed to complement their hands-on learning (described subsequently; see also the Appendix). At the end of the term students wrote a final paper in which they integrated their experiences with specific topics from the course. Finally, students completed a brief evaluation of their experiences, and agencies provided an evaluation of each of the students placed with them.

I graded the service-learning project based on three components: hours completed (minimum of 15), commitment/professional conduct (e.g., being punctual and reliable, respecting the policies set forth by the agency, maintaining confidentiality), and students’ ability to integrate their experiences with course material. Agencies reported the number of hours served and evaluated students’ conduct and contribution to the agency. I evaluated the integration with course material based primarily on students’ final papers.

Students’ Feedback

Of the 36 undergraduate students who completed the course, 32 (89%) chose the service-learning instead of the research project. Students obtained placements at 15 agencies ranging from a resource center for homeless and battered women to a home for adults with disabilities to the psychiatric ward at a local state hospital (see Table 1).

At the semester’s end all 32 students completed a 20-item evaluation form developed by the service-learning coordinator. The coordinator designed both scale-response items and open-ended questions to evaluate issues ranging from the impact of service learning on understanding course material to what students liked least about the project. Students responded to a series of statements using a 7-point scale ranging from 1 (strongly agree) to 7 (strongly disagree) to assess the three constructs included in the following analyses: academic learning process, service ethic, and agency experience. I averaged the scores across items measuring each construct to produce one score and used Cronbach’s alpha to assess the internal reliability of the items. Three items such as, “This project allowed me to apply course content to an experience outside the classroom” measured academic learning process ($\alpha = .71$); four statements such as, “Everyone has a responsibility to contribute to the improvement of their [sic] community” assessed service ethic ($\alpha = .75$); and three statements such as, “The project site provided me with appropriate supervisory assistance” measured agency experience ($\alpha = .73$).

Descriptive statistics for each of the three constructs appear in Table 2. Over 80% of the students believed that the project enhanced their learning in the course, over 90% indicated that service learning had a positive impact on their sense of responsibility, and over 90% reported positive experiences with their agencies.

Perhaps more enlightening were students’ responses to the four open-ended questions. Using qualitative research strategies (see Rubin & Rubin, 1995), I examined these responses for converging themes. The first question asked about the most meaningful aspect of service learning. Several themes were apparent. First, consistent with the goals of the project, a number of students indicated that the project enhanced their understanding of class material: “helped to understand and learn about mental illness.” Students also indicated that

### Table 1. Service-Learning Placements for General Psychology

<table>
<thead>
<tr>
<th>Placement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian Center Daycare</td>
<td>Assistance with daycare for preschool-age children from low-income backgrounds</td>
</tr>
<tr>
<td>Campus Kids</td>
<td>One-on-one mentoring for at-risk elementary school children</td>
</tr>
<tr>
<td>Center Pointe</td>
<td>Classroom support for living and recreational skills development for adults with disabilities</td>
</tr>
<tr>
<td>Crisis Nursery</td>
<td>Care of children in a temporary residence while families manage crises</td>
</tr>
<tr>
<td>Eastern State Hospital</td>
<td>Recreation supervision of adults living in a state psychiatric facility</td>
</tr>
<tr>
<td>First Call for Help</td>
<td>Phone support for individuals needing resources or who are in crisis</td>
</tr>
<tr>
<td>Habitat for Humanity</td>
<td>Support for low-income or homeless individuals seeking housing</td>
</tr>
<tr>
<td>Juvenile Detention Center</td>
<td>Classroom support and library supervision of adolescents in juvenile detention</td>
</tr>
<tr>
<td>KIDS, Inc.</td>
<td>Tutoring and recreation supervision of adolescent boys in a transitional living facility</td>
</tr>
<tr>
<td>L’Arche</td>
<td>Group home support for adults with disabilities</td>
</tr>
<tr>
<td>Sacajewa Middle School</td>
<td>Support in a resource room for children with learning and behavioral disabilities</td>
</tr>
<tr>
<td>St. Mary’s Church</td>
<td>Interaction with homeless families at an overnight shelter</td>
</tr>
<tr>
<td>Shriner’s Hospital</td>
<td>Recreational therapy with hospitalized children</td>
</tr>
<tr>
<td>Women’s Drop-In Center</td>
<td>Support for women at-risk (e.g., domestic violence, prostitution)</td>
</tr>
<tr>
<td>Youth Help Association</td>
<td>Interaction with high-risk youth at an emergency residence</td>
</tr>
</tbody>
</table>

*Long-term (e.g., more than one semester) placement.*
because of an apparent confusion about the rating scale.

Another group of students focused on the impact service learning had on them personally;

The third question asked how students’ experiences related to the course. Only 3 students (9%) indicated that the experience did not relate well to the course. In contrast, 29 students (91%) described specific connections to course material. The major theme here indicated that the experiences illustrated concepts or issues discussed in class: “I saw a lot of what we learned”; “It’s amazing to learn something and at the same time witness it happen.”

Finally, some students highlighted a heightened sense of self-confidence as a result of the project: “I truly helped some women, and this gave me a new sense of confidence.”

Another student stated that service learning “made me realize more fully the diversity of individuals in this world and the fact that each and every one of them has something worthwhile and meaningful to share.”

In sum, the students’ evaluations provided the overall impression that the service-learning project was valuable. However, from a research perspective the data can only be interpreted descriptively. The absence of a pretest makes it impossible to assess change over time, and the lack of a control group precludes any causal conclusions. For example, even though most students in the class selected service learning, students choosing this option may differ from other students in a variety of ways (e.g., more committed to service, more engaged academically). Thus, it is impossible to conclude whether service learning, itself, had an impact on the students or whether this group of students would be more likely to endorse the value of service, for example, regardless of the experience (see also Gray et al., 1996). Although causal conclusions cannot be drawn from this study, the descriptive ratings are consistent with those obtained in other descriptive and experimental studies (Batchelder & Root, 1994; Giles & Eyler, 1994; Gray et al., 1996; Markus et al., 1993). Additional research, particularly well-controlled experiments, is needed to further support the notion that service learning yields positive outcomes.

### Recommendations

In addition to providing a preliminary evaluation of service learning, students’ responses revealed some of the challenges of incorporating this kind of a project into a general psychology class. Also, having offered a service-learning option to my general psychology students for three semesters, I feel prepared to make some recommendations based on the shortcomings I have witnessed and the changes I have implemented. Readers may refer to Raupp and Cohen (1992) for additional suggestions regarding the integration of volunteer experiences with class activities.

### Project Logistics

First, and perhaps most obvious, well-organized and relevant placements are imperative. My institution is privileged to have a full-time service-learning coordinator who identified many of the initial placements from a pool of agencies indicating an interest in student volunteers. The coordinator then contacted the agencies to discuss with them the differences between volunteer service and service learning (see Cooper, 1996a; Kendall, 1991) and to discuss the specific needs of the agency and the activities in which students would be involved. Next, the coordinator and I reviewed the placements, ruling out those that I believed were less relevant to the course (e.g., sorting canned goods at the local food bank) or that required a large time commitment (e.g., Big

<table>
<thead>
<tr>
<th>Construct</th>
<th>% ≤ 3*</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic learning process</td>
<td>83.9</td>
<td>2.33</td>
<td>.87</td>
</tr>
<tr>
<td>Service ethic</td>
<td>93.5</td>
<td>1.74</td>
<td>.79</td>
</tr>
<tr>
<td>Agency experience</td>
<td>93.5</td>
<td>1.74</td>
<td>.84</td>
</tr>
</tbody>
</table>

Note. N = 31. Data from one student’s evaluation were excluded because of an apparent confusion about the rating scale.

*Students rated items measuring each construct by using a 7-point scale ranging from 1 (strongly agree) to 7 (strongly disagree); a score of 3 or less was considered a positive endorsement of the item.
Students’ Concerns

Recognizing that students are busy, many have commitments outside of school, and not all are interested in pursuing a career in the mental health field, creating reasonable requirements will enhance their experience. Indeed, when students responded to what they liked least about the project, many commented that the demands on their time were multiple and that scheduling with the agency was sometimes difficult. For these reasons, I made service learning an option (the other option was completion of a literature review focused on a particular topic of their choosing). As noted previously, most students chose service learning. Additionally, I required a minimum of 15 hr of service, flexibly scheduled, across the semester. Many students completed more hours, but they could meet the minimum expectations for the project without a huge demand on their time. For upper division courses and seminars, an increased time commitment balanced appropriately with the demands of the course may make more sense.

In addition to concerns about time constraints and scheduling difficulties, students commented that transportation was sometimes a challenge. Placements within walking distance to campus were an ideal solution, and I reminded students about the public transportation system. Additionally, after students identified their placements, I circulated a list of students at each site; made sure they knew one another; and encouraged them to car pool, walk, or ride the bus together. I also reminded students to be vigilant about their safety.

Although service learning offers ideal opportunities to discuss critical issues relevant to therapeutic relationships such as privacy, confidentiality, and termination of relationships, real risks need to be recognized. One primary concern is risk to the individual seeking services. General psychology students should not provide counseling (one-on-one tutoring or mentoring under supervision is different). Additionally, because required placements last just one semester, choosing placements in which students are likely to form close relationships with individuals (e.g., Big Brothers Big Sisters program) is not appropriate unless students are willing to make longer term commitments. It is important to make agencies aware of the time limitations; several are only on my long-term placement list. Regardless of the placement’s length, educating students about their potential impact on someone’s life and the importance of closure at termination is imperative.

Faculty must also be sensitive to students’ reactions to what they might encounter. Interacting with a child from an abusive background, talking to a woman battling a drug addiction, or getting to know a homeless family offers a powerful learning opportunity. Many of these opportunities, however, are completely foreign to the students’ range of experiences and may invoke a variety of feelings and responses. Indeed, some students expressed their frustration about not being able to do more for clients, a painful but real-life lesson about working in the mental health field. I encouraged students to talk with me or with their agency supervisors about questions or concerns. Small and large group discussions (conducted with a reminder about confidentiality) can also help students work through these experiences as can writing about them. Reflection, as noted earlier, is a critical part of the service-learning process (see Cooper, 1996b; Dunlap, 1998; Kendall, 1991; Kobrin & Mareth, 1996; Morton, 1993).

Faculty Concerns

One of the largest barriers for faculty in offering service learning is time. Adding a service-learning component does demand more time, which clearly varies depending on support provided. Institutions without a formal service-learning
coordinator may employ an individual who organizes volunteer efforts; this person could facilitate service-learning initiatives. A number of grants are available to support faculty interested in developing service-learning courses; my university’s service-learning coordinator was initially hired through a grant from Campus Compact (see Franta, 1994b; Kobrin & Mareth, 1996; Kraft & Swadener, 1994; Montrose, 1994). However, even with the support of the service-learning coordinator, my initial time investment was significant. I spent time developing and reviewing placements, creating the description and grading criteria for the project, and revising the class discussion activities to include questions about students’ service-learning experiences. I chose to keep the grading to a minimum. Instead of requiring a formal journal, I asked students to integrate their service-learning experiences into responses to discussion questions already assigned for each chapter. Students used these responses for class discussion and for exam preparation. Also, the final paper was limited in terms of length (no more than five pages) and scope. I recommended that students focus on one to three specific issues covered in the course that were most relevant to their placements and link their experiences to those issues.

A second issue concerning time is how to balance service-learning activities with the necessary coverage of the discipline. General Psychology is already a course that demands coverage of a wide range of topics. Rather than develop a separate set of service-learning activities, I integrated service-learning questions into the discussion activities already employed. In both large and small group discussions, I asked students to use their service-learning experiences to illustrate points, to provide examples, or to think more deeply about different psychological issues and concepts (see the Appendix for sample questions). For example, during some 75-min class periods small group discussions typically took place following a 20- to 30-min lecture and lasted 10 to 15 min. We would spend an another 10 to 15 min as a large group hearing one point, issue, or example from each of the small groups before moving on to additional lecture material. In general, I find that the tradeoff between the time devoted to discussion as opposed to lecture is worthwhile; the students remain engaged and by reflecting on their own experiences make connections with the material that are often more powerful than those made when I provide an example or hypothetical case.

Conclusions

Widely recognized as an effective pedagogical approach is the integration of active, hands-on, experiential learning with the intellectual rigors of the academic classroom (Cantor, 1995; Kolb, 1984). Additionally, emphasizing the education of the whole person, character as well as intellect, has received increased attention; as Anson (1993) stated: “We can move beyond the ‘me’ generations and into a new generation where the concept of ‘we’ matters, a generation just as concerned about giving out its learning as it is with taking it in” (p. 81). Indeed, this statement echos Miller’s (1969) concern that psychology students be encouraged to “give psychology away” (p. 1071). Consistent with Miller’s ideas, I suggest that fostering an appreciation for the call to service inherent in the field of psychology is an important and laudable goal for educators.

Moreover, as Franta (1994) discussed, service learning offers other benefits that may have long-term impacts on students’ educational and career paths. For example, students may continue to volunteer at their agencies or receive employment offers following the service period. A number of my students suggested that they would continue to volunteer; several were offered paid positions. Students may make connections with professionals in the mental health field who can act as mentors and provide valuable resources (e.g., letters of recommendation). Students assuredly gain experience working in the field, and this experience can serve to either confirm or challenge their educational or career choices before they must make more serious commitments. As noted previously, one student indicated that as a result of his experience he “decided not to get into psychology”; another student stated that the project confirmed her passion for the field.

Service learning may not work for every teacher or be inspiring to every student, and it may not best meet the educational objectives of every class, although creative integrations of service learning are numerous (see Ferrari & Jason, 1996; Kraft & Swadener, 1994). Additionally, further empirical work is needed to clearly assess its benefits (Bradley, 1997). However, I argue that service learning has the capacity to transform lives, to touch the heart as well as the mind, and to teach many valuable lessons beyond those that professors provide within the confines of their classrooms.

References


Appendix
Example Discussion Questions

1. Describe Maslow’s hierarchy of needs. Where do the individuals with whom you are working in your service-learning placement fit in terms of Maslow’s model?
2. Describe the central developmental task facing each individual at each stage of the life span according to Erikson’s psychosocial model. Think about individuals with whom you are working in your service-learning placement; choose one individual and describe the developmental task currently being faced by this person. Speculate, in an informed way, about how the individual will resolve the current developmental crisis.
3. Apply each of the theoretical explanations of psychopathology to a case: the boy with conduct disorder (from the film shown in class), a case from your text, or an example from your service-learning placement if relevant.
4. Changes in our social, political, and economic climates have had an impact upon treatment for psychological disorders. Describe deinstitutionalization. What would happen to the individuals with whom you have worked during your service-learning placement if the agency or institution were to no longer provide them with services?

Notes

1. I thank Debra Greenwood, Gonzaga University’s former service-learning coordinator, for developing many of the placements along with the contract and evaluation forms described and used in this article. I also thank Nancy L. Worsham, Joy Milos, Sima Thorpe, Brett Hendricks, Randolph A. Smith, and several anonymous reviewers for their comments on an earlier version of this manuscript.
2. Readers are invited to examine Gonzaga University’s Service-Learning Web site: http://www.gonzaga.edu/service/gvs/service_learning.
3. Send correspondence to Molly D. Kretchmar, Department of Psychology, 502 East Boone, Gonzaga University, Spokane, WA 99258; e-mail: kretchmar@gonzaga.edu.
Instructional Television Versus Traditional Teaching of an Introductory Psychology Course

Steven F. Bacon
Julie A. Jakovich
California State University, Bakersfield

In this study we compared the effectiveness of an introductory psychology course delivered through instructional television (ITV) with the same course taught in a traditional classroom. We compared 3 groups of undergraduates, with 83 receiving traditional classroom instruction one quarter. The next quarter, new enrollees were split into 2 groups: One received instruction in an ITV studio with the instructor \((n = 29)\); the second received televised broadcasts in a remote classroom on campus \((n = 29)\). The 3 instructional formats produced similar outcomes in performance, attrition, and attendance. Likewise, attitudes toward the course were favorable and rarely differed by format. Student attitudes toward ITV were positively affected by exposure to the experience.

Distance learning technologies hold promise for increasing access to higher education for many people who, because of scheduling conflicts (e.g., working full time, taking care of children) or remoteness from a university, have been denied it. The fulfillment of this promise, however, rests on the ability of these diverse technologies to deliver a high quality product that compares favorably with a traditional university education. One particular distance learning technology, instructional television (ITV), first appeared more than 40 years ago (Dubin & Hedley, 1969). In its simplest form, ITV involves videotaping an instructor while he or she teaches a real or simulated class in a studio and transmitting these sounds and images, live or with some delay, through television to remote learners. It is possible to add many features to this simple structure, including two-way telephone links, two-way interactive video (see Andrews, Gosse, Gaulton, & Maddigan, 1999), and computer-assisted, on-screen graphics. Initially an interesting application of a new and exciting technology in the 1950s, ITV has recently become resurgent as a way of increasing educational access. With about two thirds of U.S. homes now wired for cable (U.S. Bureau of the Census, 1998) and many educational complexes now capable of sending and receiving closed-circuit, cable, and even satellite signals, ITV represents a potentially powerful educational technology.

The most consistent finding comparing traditional classroom instruction with ITV suggests no difference between the two in student achievement (Phipps & Merisotis, 1999). Reviews reflecting this conclusion date back to Dubin and Hedley (1969), who looked at 381 studies from the 1950s and 1960s. Reviews conducted since then (Clark, 1983; Cohen, Ebeling, & Kulik, 1981; Jamison, Suppes, & Wells, 1974; Wetzel, Radtke, & Stern, 1994; Whittington, 1987; Zigerell, 1991) have continued to find no differences. Although fewer studies have investigated the effects of ITV on attendance and attrition, most have found no adverse effects on these variables (Cohen et al., 1981; Whittington, 1987). Zigerell, however, found that attrition was a greater problem for some distant learners (e.g., those who were younger and less self-motivated) than for others.

The effect of televised instruction on student attitudes toward instructors, courses, and academic experiences has also been a target of research, and, as with the more objective measures, several studies have found no differences between ITV and traditional students (Cohen et al., 1981). Some researchers have reported that students prefer a traditional classroom (Davis, 1984; Ritchie & Newby, 1989), but these studies are balanced by others suggesting students prefer televised courses over traditional ones (Zigerell, 1991). The more equivocal findings in this area may reflect the diversity of attitudes assessed by different investigators.

A recent review commissioned by the American Federation of Teachers and the National Education Association (Phipps & Merisotis, 1999) may generate new and wider interest among educators in the evaluation of distance learning technologies. This report, which received coverage in the Chronicle of Higher Education (Blumenstyk & McCollum, 1999), reviewed the literature on distance learning since 1990 and concluded that the number of empirical studies addressing the effectiveness of distance learning was surprisingly small. Phipps and Merisotis further suggested “the overall quality of the original research is questionable and renders many of the findings inconclusive” (p. 3). Although they may have overstated their conclusions, these authors reasonably challenged investigators to provide more rigorous empirical research to address this important issue.

The purpose of this study was to compare the effectiveness of an introductory psychology course taught through ITV with the same course as it has been traditionally taught. Our quasi-experimental evaluation compared remote ITV students with traditionally taught students and with students who received live lectures in an ITV studio. In addition to assessing the effect of delivery format on the objective measures of performance, attrition, and attendance, we surveyed student attitudes that might be affected by the ITV format. Based on our review of the literature, we expected no differences in performance, attrition, attendance, or course satisfaction between ITV and traditionally taught students.
Method

Participants

One hundred forty-one students participated in this study through enrollment in an introductory psychology course at California State University, Bakersfield (CSUB). Students came from three different classes. One group of 83 students, the traditional classroom students, enrolled in a Fall 1997 section of the course taught by the first author using a traditional lecture and discussion format. The other two groups of students came from two Winter 1998 sections taught by the first author.\(^1\) On the first day of instruction, winter enrollees were given the option of receiving their lectures live, in the ITV studio with the instructor, or remotely in an on-campus classroom wired to receive cable transmission from the studio. Twenty-nine students selected themselves into each of the ITV conditions: ITV live and ITV remote. Students remained in their selected condition throughout the quarter and could not attend the other section of the class.

Across the three conditions, 72.6% of the sample were women, with no significant differences in gender composition between conditions. The sample consisted of a heterogeneous mix of undergraduate majors, with those who were undeclared representing the largest number of majors (26.7%), followed by psychology (16.3%), biology (8.9%), and nursing (8.1%). To assess whether our groups were comparable in scholastic aptitude prior to taking the course, we compared three standardized aptitude scores: the nationally normed Scholastic Aptitude Test (SAT) and two California State University freshman placement exams—the Entry Level Mathematics test (ELM) and the English Placement Test (EPT).

Procedure

Students in the traditional classroom condition received lectures in a 120-seat amphitheater lecture hall. In addition to lectures, which included overhead transparencies for audiovisual support, students participated in brief small-group discussions and were encouraged to ask questions and seek clarification during lectures as needed.

Students in the ITV live condition received lectures in an ITV studio with seating for 40. As in the traditional classroom, ITV live students participated in brief small-group discussions and were encouraged to interact with the instructor as needed during lectures. The studio was equipped with a computer and an Elmo visual presenter (Model EV–400 AF, Elmo Manufacturing Corp., Chatsworth, CA) an overhead projector that can present three-dimensional stimuli onto a screen) connected to two 36-in. television monitors at the front of the classroom. The instructor created computer-generated slides, which were adaptations of the traditional classroom’s overhead transparencies, and presented these over the television monitors. An ITV technician in an adjoining room filmed the lectures by controlling the movement of three cameras mounted in the ITV studio. The technician decided which images (i.e., three views of the lecturer and classroom, computer-generated slides, visual presenter) went out over the ITV cable at any given time. Students in the ITV live condition essentially received a traditional classroom experience supplemented by the technological advantages of the ITV studio.

Students in the ITV remote condition met in an on-campus classroom away from the ITV studio. These students received lectures via ITV cable transmission to two 32-in. television monitors in the front of the classroom. A teaching assistant or nonteaching instructor proctored quizzes and exams, recorded attendance, monitored participation, maintained order, fixed technical problems, and kept small-group discussions on track. Students in this classroom were able to interact with the ITV instructor during lectures through an on-air telephone link to the ITV studio. However, technical problems and time delays made this interaction difficult, and this option was rarely exercised.

Measures

We assessed students across four broad outcome measures: attrition, attendance, academic performance, and attitudes toward the course. Attrition, attendance, and performance data were easily extracted from course grade books.

Attrition. We calculated attrition as the percentage of students who completed one quiz in the second week of classes but who failed to complete the course.

Attendance. We measured attendance in two ways. Most class sessions began with a quiz about assigned readings. We calculated one measure of attendance, quiz attendance, as the percentage of possible quizzes completed; this measure represented attendance at the beginning of each class session. Because a small number of students chose to leave class after taking their quiz, we also collected a second measure of attendance. Most sessions concluded with participation exercises, usually small-group discussions, for which students received points. The percentage of participation exercises completed, called participation attendance, represented a good estimate of attendance at the end of each class session.

Performance. We assessed overall performance in the course as the percentage of possible points that a student achieved. Students could receive points for quizzes, participation exercises, midterm examinations, an Internet assignment, and a final examination.

Attitudes and satisfaction. Students at CSUB routinely complete a standard questionnaire at the end of each course asking about the quality of the course and its instructor. To this questionnaire we added several items designed to assess students’ attitudes toward ITV. Students voluntarily and

\(^1\)A coinstructor gave lectures in 4 of the 20 class sessions in the two ITV conditions. All course evaluation materials, however, were directed toward the primary instructor.
anonymous completed the questionnaire, which used a 5-point Likert scale of item endorsement ranging from 1 (strongly agree) to 5 (strongly disagree).

Results

Comparability of Groups

We used an alpha level of .05 for all statistical tests. ANOVAs comparing the groups on scholastic aptitude showed no differences for SAT scores, \( F(2, 81) = 0.83, \text{ns}; \) ELM scores, \( F(2, 109) = 0.21, \text{ns}; \) or EPT scores, \( F(2, 97) = 2.40, \text{ns}. \) These findings suggest that our three groups were roughly equivalent for scholastic aptitude and that self-selection did not bias our sample with respect to this important variable.

Attrition, Attendance, and Performance

Attrition rates were low for all instructional formats investigated. Rates were 4.8%, 3.4%, and 3.4% for the traditional classroom, ITV live, and ITV remote conditions, respectively. Differences between groups were not significant, \( \chi^2(2, N = 141) = 0.16, \text{ns}. \) All subsequent analyses used only students who completed the course.

We conducted three separate ANOVAs to investigate the effects of instructional format on performance and the two measures of attendance. There were no differences in the percentage of total points achieved among students in the traditional classroom (\( M = 74.44, SD = 13.42 \)), the ITV live (\( M = 74.87, SD = 12.64 \)), and the ITV remote (\( M = 74.65, SD = 9.76 \)) conditions, \( F(2, 132) = 0.01, \text{ns}. \) Likewise, instructional format had no effect on quiz attendance, \( F(2, 132) = 1.38, \text{ns}. \) There were, however, differences among the three groups in participation attendance, with students in the traditional classroom attending a greater percentage of the time (\( M = 85.70, SD = 13.48 \)) than ITV live (\( M = 77.20, SD = 20.46 \)) and ITV remote students (\( M = 75.82, SD = 16.45 \)), \( F(2, 132) = 5.63, p < .01. \) A Student–Newman–Keuls follow-up test showed that the ITV live and ITV remote groups did not differ from one another in participation attendance.

Attitudes and Satisfaction

Total sample sizes for student questionnaire responses were smaller than those for performance and attendance analyses, ranging from 73 to 86, owing to different data collection procedures. Because participant responses to attitude questionnaires did not approximate a normal distribution, items were analyzed using the Kruskal–Wallis one-way ANOVA by ranks statistic, \( H, \) corrected for ties. Table 1 summarizes participants’ responses to each questionnaire item. Although the Kruskal–Wallis statistic compares groups by ranks, Table 1 shows means for ease of interpretation.

Items 1 through 5 assessed students’ attitudes toward the instructor, Items 6 through 10 assessed attitudes about the course and instructional processes, and Items 11 through 15 assessed students’ social comfort and sense of university community. Of these 15 items, only 2 discriminated between ITV remote students and those receiving live instruction: one tapping students’ perceptions that the instructor “really cared about me” and the other assessing the value of in-class activities. The last 5 items of the survey assessed students’ attitudes toward technology and its use in the classroom. Students in all three conditions held favorable opinions toward educational technologies. Interestingly, students who participated in ITV, either in the live or remote condition, held more favorable attitudes toward ITV than those who did not.

Discussion

Consistent with previous studies, there were few differences in objective outcomes between introductory psychology students who received instruction from a live lecturer and those who received ITV instruction. Level of achievement was similar across all instructional formats. In addition, we did not find the higher levels of attrition that have been found in some past comparisons with distance learners (Zigerell, 1991). We did, however, find a small but significant difference between groups for participation attendance, favoring students in the traditional classroom. Nevertheless, because the two ITV conditions were similarly affected, the effect could not have been due to live versus televised instruction. A possible explanation for this result lies with a confounded variable: The traditional class met in the morning, the two ITV sections in the late afternoon. To the extent that fatigue or home obligations (e.g., parents picking children up from school) play a role in student absence, we might expect more early departures from late afternoon classes.

For most of the attitudes assessed, including views on the instructor, the course structure, the social academic environment, and general feelings toward technology in the classroom, students held equally positive opinions across the three instructional formats. Two exceptions to this finding (Items 5 and 8) illustrate the continuing need to work toward distance learning solutions that include both high tech and “high touch” (i.e., relationship-based; Naissibt, 1984) components.

The clearest attitudinal finding in this study, mirroring an earlier observation by Wetzel et al. (1994), was that students’ feelings about ITV were positively affected by exposure.\(^3\) We did not address how exposure changed attitudes in this study. Perhaps unrealistic negative expectations were

\(^3\)An interesting complementary finding to ours is that faculty who have taught an instructional television course report enjoying the experience and wishing to do it again (Kendall & Oaks, 1992)—perhaps also an exposure effect.
**Table 1. Student Attitudes About an Introductory Psychology Course by Instructional Format**

<table>
<thead>
<tr>
<th>Item</th>
<th>Traditional</th>
<th>TV–L</th>
<th>TV–R</th>
<th>$H^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I would recommend this instructor.</td>
<td>1.33</td>
<td>1.38</td>
<td>1.32</td>
<td>0.02</td>
</tr>
<tr>
<td>2. The instructor seemed enthusiastic about teaching the subject.</td>
<td>1.31</td>
<td>1.35</td>
<td>1.29</td>
<td>0.46</td>
</tr>
<tr>
<td>3. The instructor was supportive of students and not insulting or intimidating.</td>
<td>1.31</td>
<td>1.17</td>
<td>1.50</td>
<td>2.56</td>
</tr>
<tr>
<td>4. The instructor gave clear and helpful explanations.</td>
<td>1.50</td>
<td>1.54</td>
<td>1.60</td>
<td>0.01</td>
</tr>
<tr>
<td>5. The instructor really cared about me as an individual student.</td>
<td>1.82</td>
<td>1.75</td>
<td>2.56</td>
<td>6.50</td>
</tr>
<tr>
<td>6. I would recommend this course.</td>
<td>1.36</td>
<td>1.58</td>
<td>1.37</td>
<td>0.67</td>
</tr>
<tr>
<td>7. The lectures helped me learn about psychology.</td>
<td>1.26</td>
<td>1.24</td>
<td>1.59</td>
<td>5.57</td>
</tr>
<tr>
<td>8. The in-class activities helped me learn about psychology.</td>
<td>1.55</td>
<td>1.76</td>
<td>2.19</td>
<td>8.48</td>
</tr>
<tr>
<td>9. Discussions with other students in this class helped me learn about psychology.</td>
<td>2.21</td>
<td>2.29</td>
<td>2.23</td>
<td>0.07</td>
</tr>
<tr>
<td>10. I looked forward to class sessions.</td>
<td>2.14</td>
<td>2.47</td>
<td>2.30</td>
<td>0.86</td>
</tr>
<tr>
<td>11. The class had a sense of community and I felt like a part of the group.</td>
<td>1.74</td>
<td>1.81</td>
<td>2.21</td>
<td>3.74</td>
</tr>
<tr>
<td>12. When I had questions during class, I felt comfortable asking them.</td>
<td>2.14</td>
<td>1.94</td>
<td>2.19</td>
<td>1.01</td>
</tr>
<tr>
<td>13. When I had questions or concerns outside of class, I felt comfortable talking to the instructor.</td>
<td>1.86</td>
<td>2.12</td>
<td>2.43</td>
<td>5.05</td>
</tr>
<tr>
<td>14. I was able to meet new people in this class.</td>
<td>1.83</td>
<td>1.82</td>
<td>1.95</td>
<td>0.15</td>
</tr>
<tr>
<td>15. Cal State seems like a good place to get an education.</td>
<td>1.68</td>
<td>1.88</td>
<td>1.95</td>
<td>0.63</td>
</tr>
<tr>
<td>16. I look forward to more opportunities to learn using technology (e.g., televised instruction, computer-assisted instruction, Web-based courses).</td>
<td>1.95</td>
<td>2.29</td>
<td>2.14</td>
<td>1.04</td>
</tr>
<tr>
<td>17. In general, I like trying out new technologies (e.g., computers, the Internet, Interactive television).</td>
<td>1.61</td>
<td>2.00</td>
<td>2.00</td>
<td>1.11</td>
</tr>
<tr>
<td>18. If I expect that I would learn better in a regular classroom than through instructional television.</td>
<td>1.73</td>
<td>2.57</td>
<td>2.68</td>
<td>9.63**</td>
</tr>
<tr>
<td>19. (I expect that) the instructional television format was (would be) easy to get used to.</td>
<td>2.64</td>
<td>2.00</td>
<td>1.95</td>
<td>8.73**</td>
</tr>
<tr>
<td>20. I would like to take (more) courses through instructional television.</td>
<td>3.36</td>
<td>2.25</td>
<td>2.43</td>
<td>12.81**</td>
</tr>
</tbody>
</table>

**Note.** Values are mean ratings where 1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, 5 = strongly disagree. Traditional = traditional classroom format; TV–L = instructional television format with live instructor; TV–R = instructional television format viewed from a remote classroom. $df = 2$. *Wording of item answered by TV participants. Parenthetical text added for traditional participants. *Wording of item answered by traditional participants. Parenthetical text added for TV participants.

*p < .05. **p < .01.

Our experience suggests that ITV can be an effective format for teaching an introductory psychology course. Despite the fears of some educators, our study and the literature in this area, imperfect as it may be (Phipps & Merisotis, 1999), suggest that students are not adversely affected by this alternative mode of teaching. In fact, our attitudinal data suggest that among the barriers to more frequent use of ITV may be ignorance and anxiety; students who try ITV often surprise themselves and find that they like it.

**References**


Ours was a quasi-experimental design in which ITV participants selected themselves into the live or remote condition. Although this study was mostly consistent with prior work in the area, it is important to recognize its limitations. Our experience suggests that ITV can be an effective format for teaching an introductory psychology course. Despite the fears of some educators, our study and the literature in this area, imperfect as it may be (Phipps & Merisotis, 1999), suggest that students are not adversely affected by this alternative mode of teaching. In fact, our attitudinal data suggest that among the barriers to more frequent use of ITV may be ignorance and anxiety; students who try ITV often surprise themselves and find that they like it.

Although this study was mostly consistent with prior work in the area, it is important to recognize its limitations. Ours was a quasi-experimental design in which ITV participants selected themselves into the live or remote conditions. Although we showed there were no preexisting differences in scholastic aptitude among our participants, it is possible that other uncontrolled participant variables may have influenced our results. In addition, our study did not address the cumulative effects of instructorless courses, the cost-effectiveness of distance learning approaches, or the importance of matching instructor characteristics (e.g., enthusiasm) with teaching media (e.g., ITV). We would expect our results to generalize best to other introductory psychology courses televised to distant, supervised classrooms. Departments that must offer large sections of a course but do not have the large classrooms to do so might benefit from this sort of ITV. Likewise, departments wishing to broadcast courses to distant satellite campuses with on-site teaching assistants or other classroom supervisors might be interested in these results. Our research, however, is less informative regarding home learners’ abilities to benefit from televised courses.

**References**


In this article, we discuss strategies for helping graduate students in clinical and trauma classes to cope effectively with the reactions they are likely to have while learning about violence and other trauma and working with clients who are dealing with such issues. Trauma impacts the lives of many Americans as a result of events as close to home as domestic violence, child abuse, and robbery, and as distant, for some, as Hurricane Hugo or the Columbine High School massacre. Natural and humanly caused traumatic events are common and individuals encounter such events almost daily, either personally or through the media. Mental health professionals often provide debriefing, crisis intervention, and therapy in the wake of such events.

Our goal is to discuss ways to integrate an awareness of student responses to emotionally difficult course material and to the traumatic experiences of clients in the graduate training of mental health professionals. We refer to such responses as secondary traumatic stress (STS; Figley 1995a). Our strategies and suggestions are particularly relevant to classes on trauma and violence, and the examples we discuss arise from our experiences in teaching these classes. However, it is important to note that this is not the only issue that needs to be addressed in these contexts.

In this article, we discuss strategies for helping graduate students in clinical and trauma classes to cope effectively with the reactions they are likely to have while learning about violence and other trauma and working with clients who are dealing with such issues. Trauma impacts the lives of many Americans as a result of events as close to home as domestic violence, child abuse, and robbery, and as distant, for some, as Hurricane Hugo or the Columbine High School massacre. Natural and humanly caused traumatic events are common and individuals encounter such events almost daily, either personally or through the media. Mental health professionals often provide debriefing, crisis intervention, and therapy in the wake of such events.

Our goal is to discuss ways to integrate an awareness of student responses to emotionally difficult course material and to the traumatic experiences of clients in the graduate training of mental health professionals. We refer to such responses as secondary traumatic stress (STS; Figley 1995a). Our strategies and suggestions are particularly relevant to classes on trauma and violence, and the examples we discuss arise from our experiences in teaching these classes. However, it is important to note that this is not the only issue that needs to be addressed in these contexts.

In this article, we discuss strategies for helping graduate students in clinical and trauma classes to cope effectively with the reactions they are likely to have while learning about violence and other trauma and working with clients who are dealing with such issues. Trauma impacts the lives of many Americans as a result of events as close to home as domestic violence, child abuse, and robbery, and as distant, for some, as Hurricane Hugo or the Columbine High School massacre. Natural and humanly caused traumatic events are common and individuals encounter such events almost daily, either personally or through the media. Mental health professionals often provide debriefing, crisis intervention, and therapy in the wake of such events.

Our goal is to discuss ways to integrate an awareness of student responses to emotionally difficult course material and to the traumatic experiences of clients in the graduate training of mental health professionals. We refer to such responses as secondary traumatic stress (STS; Figley 1995a). Our strategies and suggestions are particularly relevant to classes on trauma and violence, and the examples we discuss arise from our experiences in teaching these classes. However, it is important to note that this is not the only issue that needs to be addressed in these contexts.
Mathods and Techniques

Exposure to the Fields of Psychology:
Evaluation of an Introductory Psychology Project

Amanda M. Maynard, Douglas C. Maynard, and Kirsten A. Rowe
State University of New York at New Paltz

We evaluated a project designed to expose introductory psychology students to the fields of psychology. We created this project to foster student learning about the various fields of psychology, careers that psychologists have, the psychology program on their campus, and courses available at their home institution. Evaluative data included knowledge of the definitions of the fields, measured at 3 time points, as well as students’ self-reported reactions. Results indicated that students found the project interesting and would recommend it for future classes. More important, knowledge tests revealed significant increases in knowledge that persisted over a 4-week period.

Many students enrolling in an introductory psychology course do not have an appreciation of the diversity of the field of psychology or the breadth of careers available to them. Although Borden and Rajecki (2000) recommended that students be exposed to career opportunities early in their undergraduate studies, the school-to-work transition is often abrupt for graduates with a bachelor’s degree, in part because they receive little information regarding career opportunities within psychology and little guidance in how to structure their educational experience to help them obtain career goals (Borden & Rajecki, 2000; Lattall, 1980). Research suggests that psychology departments can address the need for occupational education by adding career courses to the core curriculum (Bluestein, 1977; Dillinger & Landrum, 2002; Gottlieb, 1975; Lattall, 1980; Lloyd, 2000; Zechmeister & Helkowski, 2001). Landrum, Shoemaker, and Davis (2003) reported that only about a third of psychology departments offer such a course.

Typically, career courses have targeted sophomore- and junior-level students (Gottlieb, 1975; Lattall, 1980; Lloyd, 2000) and involve presentations and lectures by individuals established in various occupations (Bluestein, 1977; Lattall, 1980; Lloyd, 2000), resulting in a relatively passive student experience. Some career courses include active learning experiences and self-study (Bluestein, 1977) and the development of a presentation or portfolio (Zechmeister & Helkowski, 2001). As a result of these courses, students reported an increase in their knowledge of careers in psychology (Lattall, 1980; Zechmeister & Helkowski, 2001). However, disadvantages of this approach include the commitment of resources that departments may not have or may not be able to sustain (Bluestein, 1977) and the constraint of increasing the number of credits within the curriculum. Offering the course as an elective does not ensure that all majors receive the information.

Career courses arose in part because the core curriculum typically uses texts offering simple definitions of the most common fields of psychology without further discussion of occupations and educational requirements needed to pursue such careers. In addition, faculty may find it difficult to find the time to add a section on careers in each course they teach. As a result, there is a need for an educational experience within the core curriculum that provides students with an understanding of the diversity of the field of psychology in concert with related career information.

The Fields of Psychology Project

We developed a project to expose students to the fields of psychology within two introductory psychology courses. We gave 63 students an assignment sheet detailing three components of the project: (a) overview of the field, (b) classroom and research opportunities on campus, and (c) career options within the field. The project included 12 fields of psychology: clinical, cognitive, consumer, counseling, developmental, exercise and sport, forensic, health, industrial/organizational, neuroscience, school, and social. First, we assigned each student to a group responsible for presenting one of the fields of psychology. Groups identified a definition of the field, journals that publish research in this field, and three sample citations of articles from this field of psychology. Next, students obtained information about classes available (and the faculty who teach them) on campus that students might seek out if interested in learning more about this field of psychology. Each group also identified faculty whose research interests fell within its area of psychology. Finally, the group researched career opportunities within this field. Students consulted the Internet, campus library, and the career advising center on campus to conduct their research. The two written products included a poster to present during a 50-min class session and a two-page summary of the poster to be duplicated and shared with all class members as a reference. Students visited each poster during the class session, asked any questions they had, and rated how interested they were in each field of psychology.

This study evaluated the success of this project in increasing knowledge and adds to the literature in this area in three ways. First, coursework in introductory psychology classes in-
cluded the fields of psychology learning experience, making the experience available to students early in their academic careers and not requiring the creation of a separate career course. Second, previous research measured learning via self-report (Lattall, 1980; Zechmeister & Helkowski, 2001). This project used a direct measure of knowledge, taken at three points during the course. Finally, the project included an opportunity to learn more about the psychology program at our university (including department faculty, ongoing research, and available classes) and careers in psychology. We expected that (a) students would demonstrate increased knowledge immediately following the project and that (b) students’ reactions to the project would be favorable.

Method

Participants

Sixty-three students from two introductory psychology classes taught by different instructors at a small public university in the Northeastern United States participated in the project. Section A included 38 students (12 men, 26 women) and Section B included 25 students (9 men, 16 women). The sample included mostly first-year students (79%). The majority of the sample was comprised of undecided students (35%) or business majors (30%). The remainder of students identified their major as psychology, education, or “other.” The sections were comparable in that both were offered in the Fall 2001 semester, used the same text, and followed roughly the same order of topics. Section B was part of a learning community; students in this section were concurrently enrolled in an introductory management course. Finally, because Section B was smaller than Section A, students in Section B worked in pairs, whereas students in Section A worked in groups of 3 or 4 students.

Measures

Knowledge. We evaluated student knowledge of the presented fields with open-ended definition questions. We focused our evaluation on understanding of the various disciplines within psychology because such understanding serves as a foundation for future learning and career exploration. To conserve class time, only 8 of 12 fields appeared on the knowledge measure (clinical, cognitive, counseling, developmental, industrial/organizational, neuroscience, school, and social). We chose these fields because most graduate students specialize in one of these areas (American Psychological Association Research Office, 2000). For each of the 8 fields, students either provided a one- or two-sentence definition or checked a box indicating that they did not know the definition.

Reactions. We generated 12 items to measure students’ immediate response to the project on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Sample items from the reactions measure included: “I enjoyed the Fields of Psychology project,” “Future Introductory Psychology classes should include the Fields of Psychology project,” and “The Fields of Psychology project broadened my knowledge of the different areas of psychology.”

Procedure

We evaluated the project by using a variation of the interrupted time series design with switching replications (Cook & Campbell, 1979). This design maximizes internal validity and interpretability of the data given nonequivalent groups. During Week 8 of the semester (Time 1), students in both sections completed the knowledge measure. During Week 9, students in Section A then presented their posters in class. In the class session immediately following (Time 2), these students completed both the knowledge and reactions measures, whereas students in Section B completed only the knowledge measure, thereby serving as a comparison group. Students in Section B then presented their posters during Week 13, and the following week (Time 3), they completed both the knowledge and reactions measures, whereas the students in Section A completed only the knowledge measure, serving as a comparison group.

Results

Because there are inherent difficulties in interpreting data for individuals who were not present at all time points and conducting split-plot analyses requires complete data for all cases, we restricted our analyses to include individuals with no missing data (N = 42; 21 students in each section). The majority of students in the subsample were first-year students (86%), with the remainder being second-year students. As with the original sample, most of these students in the subsample were either undecided (31%) or declared business (38%) majors.

Knowledge

For the knowledge data, we developed a coding scheme for the eight fields tested. We consulted introductory psychology texts for the core elements of the definition of each field and coded students’ responses as a 0 (did not know or left blank), 1 (response was only partially accurate), or 2 (answer was complete and accurate). Coding sheets included data for all three time points for each participant, but did not denote section (A or B) to avoid potential bias. Two independent raters coded the knowledge data with an overall percentage agreement of 87%. We resolved disagreements by consensus discussion.

We analyzed the data with a 3 (Time: 1, 2, 3) × 2 (Class: Section A, Section B) split-plot ANOVA with repeated measures. (See Figure 1 for the group means by time and class.) As expected, the analyses revealed a significant interaction between time and class, $F(2, 80) = 14.75, p < .001, \eta^2 = .27$. We used paired-sample t tests to interpret the interaction. For Section A, there was a significant increase in knowledge from Time 1 to Time 2, $t(20) = 5.25, p < .001$, but no significant difference in knowledge between Time 2 and Time 3, $t(20) = 1.12, p > .05$. For Section B, there was no significant difference in knowledge between Time 1 and Time 2, $t(20) =$
A significant increase in knowledge from Time 2 to Time 3, $t(20) = 8.18, p < .001$. It is important to note that knowledge gained from lecture alone would have produced identical linear trends in both sections of the course; however, the results indicated a significant increase in knowledge only immediately following the project.

Reactions

Table 1 displays descriptive statistics for each of the reaction items. The majority of students indicated that the project broadened their knowledge of psychology (83%) and exposed them to the types of careers that psychologists have (93%). Most students found the project to be interesting (79%) and indicated that the project should be included in future classes (83%). Responses to some reaction items were less favorable; fewer students agreed that the project was intellectually stimulating (51%) or helped them decide about psychology as a career (45%) or major (31%). Reaction percentages were nearly identical when we included students with some missing data ($N = 52$ or $53$, depending on the item).

Discussion

This poster project attempted to provide students with an overview of the fields of psychology, classroom and research opportunities on campus, and careers within these fields, all within the context of the introductory psychology course. Results of the evaluation indicated that the project increased students’ knowledge about various fields of psychology immediately following the poster session. Because students in each section completed the project at a different point in the semester, it is unlikely that this increase in knowledge was due to material presented in the lecture. In fact, students maintained this increase in knowledge over a 4-week interval for
tions could use the project as an extra-credit oral presentation during the semester (e.g., one per week) or, if class time is not available, as an extra-credit poster presentation displayed for students to view before and after class each week. An instructor interested in conducting the actual poster session in a larger class could also split the class into two or more smaller groups having simultaneous poster sessions running in separate classrooms.

Although students’ reactions to the project were generally positive, students did not completely agree that it was intellectually stimulating and disagreed about the extent to which the project helped them decide about psychology as a potential career or major. Instructors could address these limitations by having each group of students also interview a faculty member or professional associated with the group’s assigned field of psychology and add the information gained from the interview to the poster. This interview, which students could conduct in person, by phone, or via the Internet, might produce a higher level of intellectual stimulation and give students an insider’s perspective of a career within that field.

In conclusion, data from this study suggest that the poster project is an effective technique that allows students to explore the subfields of psychology. We have since adapted the exercise to encourage students to locate related courses, programs, and faculty, either at the home institution or at other institutions within the geographic region, particularly when few courses are available within the psychology department. With this adaptation, students have used this project to make connections between psychology and other disciplines, such as biology (neuroscience), business (industrial/organizational psychology), communication (consumer psychology), and nursing (health psychology). This active approach to learning about careers and the integration of academic disciplines may assist students in identifying possible careers earlier in their development, resulting in knowledge and skills more closely tailored to their career pursuits.

References


Notes

1. Amanda M. Maynard is now at Mount Saint Mary College.
2. We thank Laura Kim for her assistance with data entry.
3. Send correspondence to Amanda M. Maynard, Division of Social Sciences, Mount Saint Mary College, 330 Powell Avenue, Newburgh, NY 12550; e-mail: maynard@msmc.edu.
Exposure to the Fields of Psychology: Evaluation of an Introductory Psychology Project

Amanda M. Maynard, Douglas C. Maynard & Kirsten A. Rowe

Available online: 20 Nov 2009
Introductory Psychology Topics and Student Performance: Where’s the Challenge?

Andrew C. Peck and Rahan S. Ali
Pennsylvania State University, University Park

Max E. Levine
Wake Forest University

Robert L. Matchock
Pennsylvania State University, Altoona

Conventional wisdom is that some topics in introductory psychology are more difficult for students than others. Such wisdom seems reasonable given mismatches between students’ and instructors’ expectations and variations in both instructor expertise and student motivation across topical areas. Five instructors pooled students’ exam performance data to investigate topical difficulty. Data showed that students performed better on questions about social psychology, memory, consciousness, and development than questions about biopsychology and emotion. Discussion includes possible explanations and considerations for course planning.

It seems as though some topics in introductory psychology are more challenging than others. Such perceptions come from formal and informal indicators such as exam scores, student questions or requests for help, and the number of confused looks from students during lectures. It is likely that instructors draw on these perceptions to plan, prioritize, and sequence course material. Miller and Gentile (1998) found that the typical introductory psychology course lasted one semester. Our experience, which is similar to that described by Lefton (1997), is that it is quite difficult, if not impossible, to adequately cover all the chapters of a typical introductory textbook in a single term. Instructors must choose which topics to discuss, which to make students responsible for learning outside of class, and which to exclude all together. This issue is exacerbated by the trend of including interactive activities and demonstrations to bolster learning, attendance, and motivation.

Although there is some empirical work on difficulty and achievement in introductory psychology classes, the body of literature is far from extensive. Miller and Gentile (1998) found that 8% of introductory students surveyed expected the course to be easy, whereas 37% expected an intellectual challenge. Ginexi (2003) measured the difficulty of an in-
introductory course on two dimensions: how easy or difficult it was for students to comprehend their textbook and how many hours a week students estimated that they spent studying or reading for the course. Students who expected higher grades found the textbook easier to comprehend, but grade expectations were unrelated to the number of hours spent studying.

Several researchers have focused on personal factors that can affect introductory course difficulty (e.g., Beck, Rotter-Woody, & Pierce, 1991; Lange & Byrd, 2002), but minimal research has investigated topical difficulty. With respect to topical coverage, instructors believe that it is important to introduce students to different approaches psychologists take and to provide a comprehensive survey of the field, noting the importance of topics on learning, memory, physiology, abnormal psychology, personality, and social psychology (Miller & Gentile, 1998). With respect to introductory student achievement, Thompson and Zamboanga (2003, 2004) found that preclass background knowledge about psychological concepts positively predicted academic success. They also found that students performed poorly on topics that involved more domain-specific knowledge and terminology. Our project addresses topical difficulty directly and discusses some potential course planning, prioritizing, and sequencing applications.

Method

Five instructors provided exam scores from introductory psychology classes taught at the same large mid-Atlantic university. The data included 1,452 students’ answers to 968 multiple-choice questions. Exams varied in length and duration but were generally 50 to 65 questions long and lasted 50 to 75 min. Although we cannot report specific demographic data, students in our courses are representative of the university at large, with the typical student age range being 17 to 21 years. Data on incoming first-year students showed that 51% were male and 84% were White. Forty-one percent of students were in the top tenth of their graduating high school class, and 94% graduated high school with a grade point average of 3.0 or higher (http://www.budget.psu.edu/FactBook/).

Each instructor provided a unique course experience. Courses varied in terms of class size (50 to 378), classroom setting, and availability of a teaching assistant. Classes met at different times of the day using different formats (e.g., 50-min classes three times a week, 75-min classes twice a week, 150-min class once a week) and covered different topics in different sequences. Course textbooks differed and ranged in difficulty from high middle to low (see Koenig, Iliff, & Griggs, 2005). Instructors also varied in the extent to which they relied on the textbook to structure lectures and exams, asked students to rely on the textbook, and used textbook supplements.

Results

Overall, students answered 79.55% of questions correctly. To investigate the effects of topic and account for variability between instructors and their courses, we conducted a one-way between-subject ANCOVA on the percentage of correct responses for topic with instructor as a covariate. Results revealed a significant effect of topic, $F(15, 966) = 3.29, p \leq .0001$, $\eta^2 = .049$. Bonferroni comparisons indicated that questions about emotion and biopsychology were more difficult than questions about social psychology, memory, consciousness, and development ($ps < .05$). Table 1 shows the percentage of correctly answered questions by topic.

The covariate of instructor was also significant ($p \leq .0001$, $\eta^2 = .038$). Mean comparisons indicated that one instructor’s class had significantly fewer correct answers, $F(4, 966) = 12.96, p \leq .0001$, (73.59%) than the others (means ranged from 80% to 83%). The other four instructors did not significantly differ from each other. With this instructor’s data removed from the ANCOVA, topic was still significant, but the covariate of instructor was not ($p = .65$). Questions

Topics

Questions remain about the appropriate taxonomy and number of unique concepts in a typical introductory course (see Matarazzo, 1987). Some researchers have argued for a set of commonly used concepts (Quereshi, 1993; Quereshi & Sackett, 1977). More recent research (e.g., Nairn, Ellard, Scialfa, & Miller, 2003; Zechmeister & Zechmeister, 2000) has found less consistency of psychological concepts across textbooks.

We organized course content topically. Following Miller and Gentile (1998), our topics reflected related content and not necessarily the names of textbook chapters. Our topics were identical to the subdisciplines delineated by Nairn et al. (2003) except that we divided sensation and perception and combined psychological disorders and treatment/therapy. Our topics also reflected the topical structure described in Miller and Gentile (1998) except that we used holistic developmental and abnormal and clinical psychology topics. Because instructors in our sample did not cover thinking and language and because only one instructor covered industrial/organizational psychology, we excluded these topics from the study. The final list consisted of 16 topics: introduction (i.e., goals of psychology, history, and subdisciplines of the field), biopsychology (neuroscience and biological foundations), research methods, consciousness (e.g., sleep and dreams, drugs, hypnosis), sensation, perception, learning, memory, intelligence, development (child, adolescent, and adult), motivation, emotion, personality, health, abnormal psychology (disorders and their treatment), and social psychology.
instructors might consider deemphasizing biological material in introductory psychology courses. Although social psychology is an important topic that is far more prevalent in introductory psychology textbooks than questions about biopsychology and emotion were still more difficult than questions about social psychology, memory, consciousness, and development (ps < .05).

Discussion

As expected, performance varied by topic. Students performed best on social and memory questions and worst on biopsychology and emotion questions. Although the uniqueness of each classroom experience introduced much variability, we believe it adds to the external validity of our findings. Such variability does, however, preclude precise identification of causal explanations. One possible explanation is the impact of student interest and motivation (Boekaerts & Boscolo, 2002; Scheifele, 2001). When students in McKenzie and Cangemi’s (1978) study ranked their desire to learn various psychological topics, they ranked social topics at the top of the list and biopsychology topics such as the chemistry of learning and anatomy and physiology quite low. Another possible explanation involves relevant background knowledge (Glaser, 1984). Thompson and Zamboanga (2003) found that introductory students had more preclass background knowledge about themes of developmental psychology and the characteristics of REM sleep and schizophrenia and less knowledge about the James–Lange theory of emotion and the sympathetic nervous system. Posttest performance showed an advantage of prior knowledge.

Our data show that biological material can be difficult for students, suggesting that instructors reflect on the role of biological material in introductory psychology courses. Although some instructors might consider deemphasizing biological material to benefit student achievement and motivation (see McKenzie & Cangemi, 1978), we do not recommend this approach. First, introductory psychology instructors believe it is important to include physiological psychology in their courses (Miller & Gentile, 1998). Second, the American Psychological Association has recognized behavioral genetics as one of the defining facets of behavioral science (Plomin, DeFries, Craig, & McGuffin, 2003). Third, the emerging importance of biological material is reflected in one of the traditional pedagogical staples of the course—the introductory psychology textbook. Griggs, Jackson, Christopher, and Marek (1999) found that, on average, sensation and perception (9%), biology (7%), and emotion and motivation (7%) received the most coverage in introductory textbooks. Our data suggest that instructors may need to do more (e.g., examples, demonstrations, activities), not less, to bolster students’ motivation for, and learning of, biological material.

One strategy might involve devoting more class time to biological material at the expense of easier topics such as social psychology. However, social psychology is an important topic from both the instructors’ (Miller & Gentile, 1998) and students’ points of view (McKenzie & Cangemi, 1978), and exclusions could affect student motivation. Another alternative is to integrate more difficult topics into other chapters (e.g., preparing discussion of the hippocampus until the memory chapter, integrating theories of emotion into a discussion on attraction) to more evenly distribute the difficulty of the topics across the semester, connect more technical information to students’ prior knowledge (Thompson & Zamboanga, 2003), and potentially bolster motivation to learn about biological aspects of psychology. The effects of topical integration into lectures and textbook design are areas ripe for empirical investigation.

Although Laffitte (1986) found that topic order had no effect on final exam performance, student interest, or perception of course difficulty, there has been little empirical work in this area. Questions remain about a variety of topical sequences and related changes in interest, motivation, and perceived difficulty of the course at different points in the semester. Our data suggest that instructors choosing not to take an integrative approach might be able to sequence chapters to meet different goals. Introductory psychology textbooks often offer biopsychology information in the first few chapters. Although presenting biopsychology or emotion early in the course may send the message that students have to take the course material seriously, instructors might enhance learning by presenting those topics during “ fresher” points in the semester (e.g., after a break) or later in the semester using relevant prior information as scaffolding for learning. Instructors should also be able to sequence topics so that exams are more uniform in difficulty across the semester or feature certain chapters (e.g., social psychology) earlier in the course to increase motivation and self-efficacy (Lenthall & Andrews, 1983) and, ideally, student performance (Hofer & Yu, 2003).

<table>
<thead>
<tr>
<th>Topic</th>
<th>%</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social psychology</td>
<td>85.11</td>
<td>11.67</td>
<td>63</td>
</tr>
<tr>
<td>Memory</td>
<td>83.04</td>
<td>11.21</td>
<td>82</td>
</tr>
<tr>
<td>Consciousness</td>
<td>82.50</td>
<td>13.53</td>
<td>66</td>
</tr>
<tr>
<td>Development</td>
<td>81.79</td>
<td>12.32</td>
<td>99</td>
</tr>
<tr>
<td>Abnormal</td>
<td>80.50</td>
<td>14.66</td>
<td>111</td>
</tr>
<tr>
<td>Research methods</td>
<td>80.47</td>
<td>14.05</td>
<td>47</td>
</tr>
<tr>
<td>Introduction</td>
<td>80.09</td>
<td>16.49</td>
<td>44</td>
</tr>
<tr>
<td>Learning</td>
<td>79.76</td>
<td>13.16</td>
<td>89</td>
</tr>
<tr>
<td>Intelligence</td>
<td>78.79</td>
<td>17.54</td>
<td>34</td>
</tr>
<tr>
<td>Personality</td>
<td>78.75</td>
<td>15.56</td>
<td>73</td>
</tr>
<tr>
<td>Health</td>
<td>78.14</td>
<td>16.00</td>
<td>29</td>
</tr>
<tr>
<td>Perception</td>
<td>76.85</td>
<td>17.59</td>
<td>33</td>
</tr>
<tr>
<td>Sensation</td>
<td>76.57</td>
<td>11.67</td>
<td>47</td>
</tr>
<tr>
<td>Motivation</td>
<td>76.42</td>
<td>18.53</td>
<td>52</td>
</tr>
<tr>
<td>Biopsychology</td>
<td>73.42</td>
<td>13.74</td>
<td>69</td>
</tr>
<tr>
<td>Emotion</td>
<td>70.95</td>
<td>13.52</td>
<td>37</td>
</tr>
<tr>
<td>Overall</td>
<td>79.55</td>
<td>14.84</td>
<td>968</td>
</tr>
</tbody>
</table>

*Indicates that performance was significantly better than that for emotion and biopsychology (p < .05). There were no other significant differences.

Table 1. Overall Percentage of Questions Correct, Standard Deviation, and Number of Questions by Topic
Limitations

Miller and Gentile (1998) found that the typical introductory psychology course had 50 or fewer students and that many introductory courses used essay exams. Because four of the five instructors in our sample taught large-section courses and all used multiple-choice exams, our suggestions may be limited to such classes. In addition to exploring topical difficulty in more varied classroom formats, subsequent inquiry should examine what specific factors within a topic influence topic difficulty. For example, do misunderstandings about differences between the James–Lange and Cannon–Bard theories account for most of the difficulty in the emotion topic? Additional research may be needed to investigate how specific course characteristics interact with topical difficulty. Are minutes of class time on material more helpful than amount of assigned at-home reading? Is time spent lecturing more helpful than time spent discussing, showing media, or conducting demonstrations? Clearly, many important questions about topical difficulty are unexplored and offer many avenues for further exploration and discussion.

References


Notes

1. We thank Jeff Love for access to his student performance data and anonymous reviewers for helpful and insightful feedback.
2. Send correspondence to Andrew C. Peck, 126 Moore Building, University Park, PA 16803; e-mail: acp103@psu.edu.
Our study aimed to investigate which components of the guiding questions procedure used by Lawson et al. (2006) produce improved learning from educational videos and to determine whether answering guiding questions is superior to freehand note taking. We examined whether writing answers to guiding questions is superior to simply reading the guiding questions (without taking notes). Some authors (e.g., Green, 2003; Hart & Stevens, 1995) have suggested using the latter approach due to concerns that note taking may draw students’ attention away from the film. Reading guiding questions may help students identify the key concepts in the video, and this guidance may be especially important for beginning students because novices may have difficulty identifying which information is most important (e.g., Bédard & Chi, 1992).

Nevertheless, the act of taking notes (i.e., writing answers) may also be important, because it forces students to actually answer the guiding questions and take a more active approach to viewing the video. Research on student learning from lectures suggests that taking notes improves student achievement (e.g., Katayama, Shambaugh, & Doctor, 2005; Titsworth &
Although taking notes may be important, we doubt it would work as well without guiding questions. Kreiner (1997) found that taking notes on an educational video without guiding questions did not improve students’ performance compared to a control group who took no notes. Moreover, research on lecture note taking suggests that (a) students take incomplete notes, recording less than 40% of the important points; and (b) providing students with an outline or organizational cues to facilitate note taking increases the number of important ideas students record and improves achievement (e.g., Kiewra, Benton, Kim, Risch, & Christensen, 1995; Titsworth & Kiewra, 2004). Likewise, giving students guiding questions to answer while watching an educational video may improve their achievement by helping them select, encode, and record the important information in a video.

We showed introductory psychology students a video about intelligence and testing (Zimbardo, 2001b). We gave four types of instructions: watch the video without taking notes, take notes, think about guiding questions while watching the video, or write answers to the guiding questions. Afterward, students took a quiz on the video information. Half of the quiz items were related and half were unrelated to the guiding questions. We hypothesized that students who wrote answers to the guiding questions would outperform the other groups on the quiz items related to the guiding questions but not on the other items.

Method

Participants

A total of 116 students enrolled in one of six sections of introductory psychology participated; we dropped 3 students because they did not follow instructions. Thus, 113 students (53 men and 60 women) comprised the final group of participants. There were 30 participants in the control condition, 27 in the note-taking condition, and 28 in each of the two guiding-questions conditions. The average age of participants was 19.50 (SD = 4.24).

Procedure

Approximately 9 to 10 weeks into a 15-week semester, three instructors (the authors)—each teaching two sections—showed the “Testing and Intelligence” video from the Discovering Psychology series (Zimbardo, 2001b). We showed the video at the beginning of class, before teaching any material for the intelligence segment of the course. Prior to showing the video, we distributed instruction sheets that informed students we were testing the effectiveness of several different techniques for viewing educational videos. The sheet explained that some students would receive different instructions than other students and that they should follow their instructions and not be concerned if “other students are doing something that you are not doing.” The instruction sheet also requested each student’s age, sex, and name.

We randomly assigned students within each course section to one of four conditions. The instruction sheet directed students to (a) watch the video without taking notes (control condition), (b) take notes on the video using the back of the instruction sheet (notes condition), (c) read the guiding questions on the back of their instruction sheet and “keep the questions in mind while watching the video . . . [but] do not write answers to these questions” (guiding-questions condition), or (d) read the guiding questions and write answers while watching the video (guiding-questions-notes condition). In the latter two conditions, the back of the instruction sheet contained seven open-ended questions (e.g., “How does Robert Sternberg define intelligence?”) about factual information in the video. As suggested by Lawson et al. (2006), the questions closely matched the wording used by the narrator to introduce a topic before he discussed it.

Immediately after showing the video, we collected students’ instruction sheets and gave them a 14-question multiple-choice quiz that did not count toward their course grade. Prior to this point, we did not inform students about the quiz. The quiz consisted of seven questions on video information related to the guiding questions and seven questions on video information unrelated to the guiding questions. We included the latter questions to determine if the guiding questions interfered with students’ ability to learn video information unrelated to the guiding questions. We alternated the guiding-questions-related and guiding-questions-unrelated items throughout the test. In a later class, we gave students the answers to the quiz questions and to the guiding questions so that the students in all four groups received the same information.

Results

We calculated the percentage of correct answers for each student on guiding-questions-related items and
Table 1. Mean Percentage Correct on Quiz Questions Related and Unrelated to the Guiding Questions as a Function of Experimental Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Related</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Control</td>
<td>57.62</td>
<td>23.27</td>
</tr>
<tr>
<td>Notes</td>
<td>65.61</td>
<td>15.98</td>
</tr>
<tr>
<td>Guiding-questions</td>
<td>68.37</td>
<td>22.12</td>
</tr>
<tr>
<td>Guiding-questions-notes</td>
<td>85.20</td>
<td>13.74</td>
</tr>
</tbody>
</table>

An ANOVA was performed on guiding-questions-unrelated items and performed a 2 (question type: related vs. unrelated) × 4 (condition: control vs. notes vs. guiding-questions vs. guiding-questions-notes) mixed-design ANOVA on the data. The results revealed the expected significant interaction, $F(3, 109) = 9.03$, $p < .001$, $\eta^2 = .20$.

To examine the nature of the interaction, we performed a simple main effects analysis to test the effect of experimental condition for each question type. For the test items related to the guiding questions, there was a significant effect of experimental condition, $F(3, 109) = 10.33$, $p < .001$, $\eta^2 = .22$ (see Table 1 for cell means). A Tukey’s honestly significant difference (HSD) test revealed that students in the guiding-questions-notes condition scored significantly higher than did students in each of the other three conditions, all $p$s < .009. None of the other mean differences were statistically significant, all $p$s > .15. For the test items unrelated to the guiding questions, there was no significant effect of experimental condition, $F(3, 109) = 1.12$, $p = .35$, $\eta^2 = .03$.

Discussion

The results of this study replicate and extend the findings of Lawson et al. (2006) with different instructors, students, guiding questions, and a different video; students who answered guiding questions while watching an educational video had the highest quiz performance on items related to the guiding questions. The average score among students in the control condition corresponded to an F on our grading scale, whereas the average score of those in the guiding-questions-notes condition corresponded to a B. We believe this is a valuable, if not dramatic, improvement in their achievement.

We were also encouraged by the fact that the guiding questions did not significantly impair students’ learning from portions of video that were unrelated to the guiding questions. One might suspect that students would pay attention to only those portions of the video that were related to the guiding questions. However, our results showed that the performance of students who had guiding questions did not differ significantly from other groups on the test questions unrelated to the guiding questions; in fact, their performance closely matched that of the control group. Of course, this result also suggests that the effect of writing answers to guiding questions does not generalize beyond the topics covered by those questions. Thus, instructors should carefully consider which video topics to cover in the guiding questions.

Although guiding questions might help students identify key concepts in a video, our results showed that it is not enough to tell students to think about the questions while watching the video; they also need to write answers. Asking students to write answers to guiding questions encourages them to take a more active approach to learning and might also make them feel more accountable for learning the information. As Katayama et al. (2005) stated, “the process of taking . . . notes is an active process that increases the internal encoding of information being learned” (p. 131). Contrary to some authors’ (e.g., Green, 2003; Hart & Stevens, 1995) proscriptions against note taking during videos, we found that (a) taking freehand notes did not hurt students’ performance and (b) taking notes to answer guiding questions substantially improved their performance.

Nevertheless, there is some debate about whether the act of taking notes itself enhances learning or whether reviewing one’s notes is essential for learning (e.g., Baker & Lombardi, 1985). Because we had students take a quiz before they had an opportunity to review their notes, our results (in the guiding-questions conditions) are consistent with the notion that the act of taking notes increases learning by enhancing encoding of the information; other researchers have drawn similar conclusions (e.g., Baker & Lombardi, 1985; Katayama et al., 2005). Although taking notes may enhance encoding, it is also important to provide students with some guidance on the key concepts that they need to record (e.g., Kiewra et al., 1995; Titsworth & Kiewra, 2004). Our students who took freehand notes did not outperform the control group, but those who wrote answers to guiding questions outperformed the other groups (i.e., on items related to the guiding questions).
Further research is needed to determine whether the effects we observed generalize to situations in which students (a) expect to receive graded quizzes on videos, (b) have time to review their notes before taking a quiz, (c) receive additional in-class coverage of the video topics, or (d) are tested several weeks after viewing the video. The immediate benefits from guiding questions that we observed may be strengthened if students encounter the material again in class or while reviewing their notes or if student motivation is increased by knowing a quiz will be graded. Delaying the quiz several weeks may decrease students’ performance, especially if there is no additional review of the material; but if answering guiding questions improves encoding of the material, students who answered those questions may still outperform those who did not.

In short, providing students with guiding questions to answer while watching an educational video appears to be an easy, effective way to enhance students’ learning. As Lawson et al. (2006) noted, instructors can expand this technique by having students compare their answers in small groups after viewing the video; each small group could then report an answer to the entire class. This practice may encourage active learning, provide additional review of the material, and help students feel more accountable for answering the questions.

References


Notes

1. We thank Liz Barkley for entering our data into SPSS.

2. We presented an earlier version of this article at the 13th Annual APS-STP Teaching Institute, New York, May 2006.

3. Send correspondence, including requests for the full set of guiding questions, to Timothy J. Lawson, Psychology Program, College of Mount St. Joseph, Cincinnati, OH 45233; e-mail: Tim_Lawson@mail.msj.edu.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
Classic Articles as Primary Source Reading in Introductory Psychology

Richard A. Griggs  
University of Florida

Sherri L. Jackson  
Jacksonville University

Given the frequent use of primary source readings by introductory psychology teachers, especially at liberal arts colleges, we compiled the citation frequencies of classic articles in introductory textbooks to help teachers select readings for their courses. Using a random sample of 24 of the most recent introductory texts (copyright dates 2004–2006), we checked for citation of the 40 classic studies that Gorenflo and McConnell (1991) identified as cited in a simple majority of textbooks in the late 1980s. We found that 22 of these articles still met this citation criterion. We discuss the general content of these articles and ways to supplement them to comprise a sufficiently diverse set of possible readings for the introductory course.

Primary source readings are a valuable component of the undergraduate psychology curriculum, especially at liberal arts colleges (Oldenburg, 2005). Oldenburg (2005) found that about 71% of teachers at liberal arts colleges assign primary source material. Although the assignment of such material was more prevalent in intermediate and upper division psychology courses, almost half of the introductory psychology teachers (46%) required such reading, especially journal articles.

Given this substantial use of journal articles in the introductory course, information about which classic articles appear most often in introductory textbooks should help teachers select readings for their courses. This information would seem especially valuable to relatively inexperienced introductory teachers (graduate students and assistant professors), a large percentage of instructors teaching the course (Griggs, Lange, & Meyer, 1988; Miller & Gentile, 1998). Gorenflo and McConnell (1991) provided citation information for 24 introductory textbooks published between 1985 and 1989 and compiled a reader of the most commonly cited classic articles (McConnell & Gorenflo, 1989) based on their findings. Whereas individual textbook authors would have their own biases in choosing which articles to include in their texts, selecting the most common articles from a large sample of textbooks compensates for the idiosyncrasies within any single text. This procedure also avoids the rather arbitrary criteria used by editors for inclusion in readers (Shima, 1977).

Given that Gorenflo and McConnell (1991) assessed introductory texts dating back as far as 21 years ago, the citation frequencies for many of these classic articles may have changed. Current introductory textbooks may no longer discuss many of these classic articles because of market pressure for citation currency (Griggs, Proctor, & Cook, 2004; Weiten & Wight, 1992). For example, Griggs et al. (2004) found that 70% of the book citations in introductory textbooks had copyrights from 1980 onward. Authors may drop classic studies to make room to cover newer research. To check this possibility, we examined the citation status of the classic journal articles identified by Gorenflo and McConnell (1991) in a sample of current introductory textbooks. Current citation frequencies will better inform introductory teachers considering such articles as primary source reading for their courses.

Method

To sample the most recent introductory textbooks, we obtained the list of textbooks included in the 2003 to 2006 compendium of introductory psychology textbooks that was in preparation at the time of this study (C. S. Koenig, personal communication, February 7, 2006). The compendium lists, describes, and compares all of the introductory psychology textbooks published within the latest 4-year period (available online at http://www.lemoyne.edu/OTRP/introttexts.html). We excluded briefer versions of regular-length introductory textbooks because their reference lists are essentially
subsets of the reference lists for their companion textbooks. From the set of 39 regular-length texts listed, we excluded those with a copyright date earlier than 2004 (5 texts) to ensure that we sampled from the most recently published texts. We also excluded any brief texts (2 texts) because they did not include all of the chapter topics in which citations of classic studies might appear. These exclusions left 32 texts from which we randomly sampled 24 to match the size of the Gorenflo and McConnell (1991) sample, allowing the direct comparison of citation frequencies in the two studies.

The 24 textbooks included in the sample are listed in Appendix A. The sample included at least 60% of the texts at each of the five levels of text difficulty used to structure the compendium. The edition number of the 24 sampled textbooks ranged from 1 to 17 (M and Mdn = 6). There were 13 texts with a copyright date of 2004, 6 with a 2005 copyright date, and 5 with a 2006 copyright date. These copyright dates thus ranged from 15 to 19 years since the earliest text copyright date in Gorenflo and McConnell’s (1991) text sample. Eight of the 24 sampled texts also appeared in the Gorenflo and McConnell sample (given in McConnell & Gorenflo, 1989), but in much earlier editions. We identify these eight texts in Appendix A.

We then identified all articles cited in a simple majority of the textbooks (i.e., more than 12) from the Gorenflo and McConnell (1991) study (see Appendix B). We tabulated the citation frequency in our sample of textbooks for each of the 40 articles.

Results and Discussion

Table 1 provides the citation frequencies for each of the 40 articles in this study and in the Gorenflo and McConnell (1991) study. As would be expected given the currency bias in citation, the majority of articles (25 of 40) decreased in frequency. The largest decreases were for the Rodin (1981) and the Bem and Allen (1974) articles, minus 17 and 19, respectively. The frequency of 4 articles remained the same, and 11 articles increased in frequency. The frequencies of 9 of these 11 articles increased by only 1 or 2, but the Brown and Kulik (1977) article increased by 3 and the Garcia and Koelling (1966) paper by 5. None of the articles was cited in all 24 textbooks. Two articles, Miller (1956) and Sperling (1960), appeared in 23 textbooks. Twenty-two of the 40 articles were still cited in 13 or more of the 24 textbooks, giving the introductory teacher a good selection of classic articles for primary source assignment. Although clearly biased toward learning, memory, and social psychology (14 of the 22 articles are from these three chapters), this set of articles spans most of the different chapter topics in introductory textbooks. This bias seems reasonable given that learning, memory, and social psychology are the first, second, and sixth most frequently assigned topics in the introductory course (Miller & Gentile, 1998). However, it is more problematic that there are no articles from developmental psychology, personality, or psychobiology. The finding that there are no articles from the developmental and personality chapters is likely due to the fact that much of the classic work by the relevant major figures in these areas (e.g., Erikson, Freud, Maslow, Piaget) was published mainly in book rather than journal format. Introductory teachers interested in including some classic reading from these areas (and others) can find relevant data on classic book citation in Griggs et al. (2004).

With respect to classic psychobiology articles, none appeared in a simple majority of the textbooks in Gorenflo and McConnell (1991). To provide introductory teachers with additional information on classic psychobiology articles, we examined the current citation frequencies for four articles in this area that appeared in 10 to 12 textbooks in the Gorenflo and McConnell study. These four articles were Gazzaniga (1983), Geschwind (1979), Lashley (1950), and Olds and Milner (1954). Gorenflo and McConnell found the frequencies of these four articles to be 10, 11, 11, and 12, respectively. We found that the Olds and Milner paper was commonly cited in current textbooks, with a frequency of 13. The highest frequency for any of the other three articles was 8.

The Olds and Milner (1954) article, the other 22 articles identified in this study as still commonly cited in contemporary introductory textbooks, and the most frequently cited classic books identified by Griggs et al. (2004) comprise a set of classic works of suf-
### Table 1. Citation Frequencies for the 40 Articles

<table>
<thead>
<tr>
<th>Articlea</th>
<th>Citation Frequencyb</th>
<th>Citation Frequency (Gorenflo &amp; McConnell, 1991)b,c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miller (1956)</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>Sperling (1960)</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>Schachter &amp; Singer (1962)</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Peterson &amp; Peterson (1959)</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Watson &amp; Rayner (1920)</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Festinger &amp; Carlsmith (1959)</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Milgram (1963)</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Holmes &amp; Rahe (1967)</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>Craik &amp; Lockhart (1972)</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>Garcia &amp; Koelling (1966)</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Gibson &amp; Walk (1960)</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Tolman &amp; Honzik (1930)</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Gardner &amp; Gardner (1969)</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Brown &amp; Kulik (1977)</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Rosenhan (1973)</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Zajonc (1968)</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Cannon (1927)</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>Brelan &amp; Brelan (1961)</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Darley &amp; Latané (1968)</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Asch (1956)</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Hobson &amp; McCarley (1977)</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>Duncker (1945)</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Brown &amp; McNeil (1966)</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>Eysenck (1952)</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Jensen (1969)</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Bandura et al. (1963)</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Scarr &amp; Weinberg (1976)</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>Ekman et al. (1983)</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Bower (1981)</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Smith &amp; Glass (1977)</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Dutton &amp; Aron (1974)</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Loftus &amp; Loftus (1980)</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Walster et al. (1966)</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Zajonc (1984)</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Holmes (1984)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Schachter (1971)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Skeels (1966)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Marshall &amp; Zimbardo (1979)</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Bern &amp; Allen (1974)</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>Rodin (1981)</td>
<td>3</td>
<td>17</td>
</tr>
</tbody>
</table>

The articles appear in order of descending citation frequency in this study. When article frequencies are equal, the articles appear in descending order based on the citation data in Gorenflo and McConnell (1991). Complete references for the 40 articles appear in Appendix B. \(^N = 24\). There was a discrepancy between the data reported in Gorenflo and McConnell (1991) and McConnell and Gorenflo (1989). McConnell and Gorenflo reported that Asch (1951) was cited in 18 of 24 textbooks. Gorenflo and McConnell, however, did not even mention this article. Given this discrepancy, we checked the 24 textbooks in this study for citation of this Asch article. This article was cited in only 11 textbooks.

Referring to the study of psychology. However, we recommend that introductory teachers exercise some caution with respect to how many and which of these works they assign. Some of these works may prove too demanding for the introductory student. There is essentially no
References


Appendix A

References for the 24 Introductory Textbooks Used in Study


Appendix B

References for 40 Articles Cited in the Majority of the Introductory Textbooks in Gorenflo and McConnell (1991)


*These eight textbooks (in much earlier editions) also appeared in the Gorenflo and McConnell (1991) sample.


Notes

1. We thank three anonymous referees for their comments on an earlier version of this article.
2. Send correspondence to Richard A. Griggs, Department of Psychology, PO Box 112250, University of Florida, Gainesville, FL 32611; e-mail: rgriggs@ufl.edu.
TOPICAL ARTICLES

The Effect of Refuting Misconceptions in the Introductory Psychology Class

Patricia Kowalski and Annette Kujawski Taylor
University of San Diego

Students often come into the introductory psychology course with many misconceptions and leave with most of them intact. Borrowing from other disciplines, we set out to determine whether refutational lecture and text are effective in dispelling student misconceptions. These approaches first activate a misconception and then immediately counter it with correct information. We tested students' knowledge of 45 common misconceptions and then taught the course with lecture and readings of a refutational or standard format or did not cover the information at all. Students showed significant changes in their beliefs when we used refutational approaches, suggesting refutational pedagogies are best for changing students' misconceptions.

“Most people use only 10% of their brain power.” “Mozart's music can increase infants' intelligence.” “The right half of the brain is the creative side.” Such claims are popular despite their lack of supporting evidence. Although unfortunate, it is at least understandable that the public believes such claims, given the predominance of unsubstantiated claims in the media (Lilienfeld, Lynn, & Lohr, 2004). As teachers of introductory psychology, we hope that presenting information supported by evidence will decrease psychological misconceptions. Sadly, this belief in our effectiveness in reducing student misconceptions is itself a misconception. Over 80 years of research suggests not only that students come into our classes with a wide variety of misconceptions, but that they also leave with their erroneous beliefs intact (Garrett & Fisher, 1926; McKeachie, 1960; Taylor & Kowalski, 2004; Vaughan, 1977).

Why do instructors see so little change in students’ misconceptions following the introductory course? Some researchers have argued that this observation is an artifact of measurement (Barnett, 1986; Gardner & Dalsing, 1986; Griggs & Ransdell, 1987). Griggs and Ransdell (1987), for example, noted that misconception tests using a true–false format often include items that are not completely false or are not addressed in introductory psychology textbooks. In response, researchers have measured a variety of misconceptions, often taken directly from textbook resources (e.g., Miller, Wozniak, Rust, Miller, & Slezak, 1996), and have developed diverse instruments to measure student misconceptions (e.g., Gardner & Dalsing, 1986; McCutcheon, 1991; Taylor & Kowalski, 2004). Despite this diversity of instruments, researchers continue to find that students hold many misconceptions that are difficult to change.

Although it is clear that students hold misconceptions, it is less clear how to reduce those misconceptions. Neither McKeachie (1960) nor Vaughan (1977) found evidence that the introductory course promoted general thinking skills that allowed students to apply and generalize their learning. Even when instructors address topics in class, students appear not to see the relevance of the discussion to their current beliefs. The few misconceptions that changed in Vaughan's study tended to be those that were directly refuted in the readings. As a result, Vaughan suggested that instructors should identify student misconceptions, then directly refute them through readings and lectures.
Literature on conceptual change learning specifically addresses the value of targeting and refuting misconceptions as a way of altering false beliefs (Guzzetti, Snyder, Glass, & Gamas, 1993; Posner, Strike, Hewson, & Gertzog, 1982). According to this literature, when correct prior beliefs provide a foundation for information, learning is facilitated, but when prior beliefs contradict new information, the prior beliefs hamper learning. For change to occur, individuals must become dissatisfied with their prior belief and find new conceptions that are intelligible, plausible, and useful. Teaching for conceptual change, therefore, often involves engaging students in activities or demonstrations designed to create cognitive conflict between their prior knowledge and the information to be learned. In a meta-analysis of instructional strategies, Guzzetti et al. (1993) showed that such conceptual change instruction, when compared to standard instruction, had a greater impact on student misconceptions.

Some research on students’ misconceptions in psychology confirms the value of classroom activities in decreasing beliefs in individual psychological misconceptions (e.g., 10% brain use; Higbee & Clay, 1998). However, others have found activities only slightly effective (e.g., “extramission” as the basis for vision; Winer, Cottrell, Gregg, Fournier, & Bica, 2002). Outside psychology, educational research (e.g., Bransford, Brown, & Cocking, 2000) suggests that despite the importance of activities, the lecture plays a critical role in promoting meaningful conceptual change because it helps students to organize and attend to the new information.

Few studies have addressed the role that various methods of presenting information play in changing psychological misconceptions. In one such study, Miller et al. (1996) identified a set of misconceptions and then addressed them in class by text, lecture, both, or neither. The study demonstrated the value of calling students’ attention to their misconceptions and explicitly directing their attention to the evidence, but cautioned against reliance on reading alone to make the connections obvious to students.

Reading researchers have specifically explored whether text can facilitate change in student misconceptions. Guzzetti (2000), for example, noted the value of a particular kind of text, referred to as “refutational text” (p. 90), in dispelling misconceptions. Refutational text directly addresses a common misconception, and then refutes the misconception by presenting evidence supporting the correct information. Hynd, Alvermann, and Qian (1994) provided evidence that refutational text is superior to standard text (which discusses the concepts but does not activate the misconception). Guzzetti suggested that refutational text is superior to standard text in reducing misconceptions because it makes explicit the incongruence between the students’ current thinking and the textbook, and students are therefore less likely to ignore the new information.

Similar to research on conceptual change, studies in the reading literature have found that refutational text alone might not be enough to change some students’ misconceptions. Marshall (1989), for example, found that a condition in which students observed a demonstration and then read refutational text produced greater change than a condition in which students read first and then observed the demonstration. Both conditions, however, were superior to conditions in which students simply read refutational text or observed demonstrations. Other research also found that individuals might need guidance from the teacher to direct their attention to arguments that refute non-scientific conceptions (Alvermann, Hynd, & Qian, 1995; Guzzetti et al., 1993).

From this research on conceptual change and reading, it appears that although changing students’ misconceptions is difficult, change can occur when instructors confront misconceptions directly with refutational text and direct student attention to the refutation in class. No study has addressed the use of refutational text and lecture in reducing student misconceptions in the introductory psychology class. In this study we used an in-class design that allowed us to compare students’ change in misconceptions in introductory psychology in response to various combinations of pedagogies. We asked the following research questions: (a) Are students more likely to change their misconceptions with refutational text and lecture, compared with standard text and lecture, or with no coverage at all? (b) Do students’ misconceptions change when they read refutational text alone, or does change require the refutational lecture?

Method

Participants

The participants were 65 introductory psychology students enrolled in sections taught by the authors during the fall semester of 2006 at the University of San Diego. The sample was 80% women, and the average age was 18. Participants completed the study as
part of their course; however, we included data only from students who provided written consent (87% of all enrolled students).

Materials

Psychological Information Questionnaire. We used the true–false format, including both true and false items, in contrast to other questionnaires, on which all of the items appear as false to be correct. The questionnaire contained 100 true–false items assessing students’ knowledge of psychological information. Fifty-five items reflected facts normally covered in an introductory psychology course (i.e., “Psychology is defined as the science of behavior and mental processes”). We included these items to minimize demand characteristics and to prevent a response set bias (for false). We then randomly embedded the 45 misconceptions items, which are the focus of this article, selecting items from previous instruments (e.g., “A schizophrenic is someone with a split personality”; Vaughan, 1977), from the popular literature (e.g., “Mozart’s music increases infant intelligence”), and from an instructor’s manual (Bolt, 2007; e.g., “Human intuition is remarkably accurate and free from error”).

Readings. The course text was Melucci’s (2004) Psychology: The Easy Way. This concise text allowed us to control the type and amount of information to which we exposed the students. We supplemented the text with 38 readings, 17 directly related to specific misconceptions assessed on the Psychological Information Questionnaire. The remaining 21 readings related to other information and masked the intent of the 17 readings targeting misconceptions. Reading length ranged from 1 to 15 pages. Sources for readings included chapters from books (e.g., Stanovich, 1998), periodicals (Wallis, 2004), and Internet sources (e.g., “Young Brains on Alcohol,” 2002). Readings addressed claims in either a standard or a refutational manner, as described by Guzzetti (2000). Table 1 provides examples of refutational and standard text.

Course Design

We designed the course to allow us to examine the different methods of addressing misconceptions in lecture and in readings. We covered items from the Psychological Information Questionnaire in either a refutational lecture, a standard lecture, or not at all. We also provided readings with refutational text, standard text, or none at all. Because of constraints created by the naturalistic setting of this study, we assessed five combinations of coverage, each with nine items, in a within-subjects design: (a) refutational lecture and refutational reading (R/R), (b) refutational lecture and no reading (R/N), (c) standard lecture and standard reading (S/S), (d) no lecture and refutational reading (N/R), and (e) no lecture and no reading (N/N).

Procedure

Participants completed the Psychological Information Questionnaire as a pretest during the first class. During the semester, we assigned readings in the Melucci (2004) text as well as additional articles. We covered item content in class with refutational

Table 1. Examples of Refutational and Nonrefutational Texts

<table>
<thead>
<tr>
<th>Refutational Text</th>
<th>Nonrefutational Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many introductory psychology books have a figure in the chapter on the senses that shows how sensitivity to salty, sour, sweet, and bitter tastes vary over the surface of the tongue. They show an area at the tip that is sensitive to sweet, two areas a little to either side that are sensitive to salty, two farther back along either side that are sensitive to sour, and a single area in the back of the tongue that is sensitive to bitter. Only one problem: It isn’t so. It is true that sensitivity to the different taste qualities varies somewhat over the tongue, and the areas listed are for the most part the ones most sensitive to the various tastes. But with the exception of the middle of the tongue, which is totally insensitive to any taste, all areas of the tongue are sensitive to all taste qualities. (McBurney, 1996, p. 21)</td>
<td>It’s generally (but not universally) agreed that there are four primary tastes: sweet, sour, bitter, and salty. Sensitivity to these tastes is distributed somewhat unevenly across the tongue, but the variations in sensitivity are quite small and highly complicated. Although most taste cells respond to more than one of the primary tastes, they typically respond best to one. Perceptions of taste quality appear to depend on complex patterns of neural activity initiated by taste receptors. (Weiten, 2004, p. 159)</td>
</tr>
</tbody>
</table>
lecture, standard lecture, or not at all. Exams included multiple-choice or short-answer items over the readings and lectures. On the last day of class, students again completed the Psychological Information Questionnaire in class.

Results

For all analyses we used \( p < .05 \) as the criterion for determining significance of findings. We calculated percentage correct on the 45 misconception items at both pretest and posttest, as well as for the subsets of items addressed by each of the five methods of coverage. Students averaged 30.09\% (SD = 8.37) correct on the pretest, and 64.41\% (SD = 12.44) correct at posttest.

In addition to percentage correct, for the misconceptions items we calculated average normalized gain scores according to the method described by Hake (2002, 2005) and implemented by Coletta and Phillips (2005) for evaluating pedagogical techniques in physics and by Grossman (2005) in psychology. This technique defines normalized gain \(<g>\) as actual gain divided by maximum possible gain (Hake, 2002). We computed each student’s individual normalized gain and then averaged these gain scores across the group of students. This procedure resulted in average normalized gain scores ranging from 0.00 to 1.00, with 1.00 indicating a greater overall average normalized gain. The advantage of using average normalized gain scores is that the final evaluation of posttest scores takes into account the level of performance at pretest.

The initial analysis involved a one-way repeated measures ANOVA assessing the effect of the five different coverage methods on \(<g>\). This overall ANOVA was significant, \( F(4, 61) = 110.42, p < .001, \eta^2 = .88 \). We then addressed the first two research questions by conducting tests of within-subjects contrasts, controlling \( \alpha \) across all contrasts.

Question 1

Are students more likely to change their misconceptions with refutational text and lecture, compared with standard text and lecture, or with no coverage at all? This question involved comparing R/R to S/S and R/R to N/N. The results, with standard deviations noted in parentheses following means, showed a greater \(<g>\) for R/R = .72 (24), compared with S/S = .37 (28), \( F(1, 64) = 65.93, p < .001, \eta^2 = .51 \), and a greater \(<g>\) for R/R compared with N/N = .19 (25), \( F(1, 64) = 271.56, p < .001, \eta^2 = .81 \). In addition, the relative gain comparing S/S to N/N was also significant, \( F(1, 64) = 17.16, p < .001, \eta^2 = .21 \). As Figure 1 more clearly shows, these results suggest little gain in scientific understanding for claims not covered in class. Students' misconceptions about psychological concepts decreased with the standard approach to covering topics, when covered both in text and in lecture. Further, the technique of directly refuting misconceptions in text and lecture was superior to the standard approach.

Question 2

Do students’ misconceptions change when they read refutational text alone or does change require the refutational lecture? This question involved comparing R/R to R/N and comparing R/R to N/R. Test of within-subjects contrasts showed no difference in \(<g>\) with R/R = .72 (24), compared to R/N = .68 (23). In contrast, we did find a significant difference in gain when comparing R/R to N/R, \(<g> = .34 (34), F(1, 64) = 88.12, p < .001, \eta^2 = .58 \). This finding, depicted in Figure 2, suggests that not only is the refutational lecture important in producing gains in students’ understanding of psychological misconceptions, but also that refutational text might add little to the effect.
Discussion

The purpose of this study was to determine whether a specific method of addressing psychological misconceptions in course text and lecture would produce gains in student understanding. Despite high levels of initial misconceptions, the refutational method for presenting new information influenced whether students developed accurate scientific knowledge. As others (e.g., Guzzetti, 2000) noted, we found that directly refuting misconceptions in class seemed to be particularly important. It could be that the lecture acts as an organizing tool, helping students to first activate their prior, incorrect knowledge and then allowing for storage of a second, stronger bit of information that contains the correct knowledge.

Although it is beyond the scope of this article, it is probable that both pieces of information coexist in memory, with the misconception now labeled as "correct" (see Reed, 2007, for a discussion of replacing vs. updating schemata). The refutational lecture did result in greater understanding of the typically misunderstood concept compared with the standard lecture and text. In fact, although several past studies found only a 5% to 7% change in students' psychological misconceptions following the introductory class (Gardner & Dalsing, 1986; McKeachie, 1960; Vaughan, 1977), we found a 34.3% change. When we presented information in a refutational manner, students showed the greatest net change, improving performance by 53.7%. In addition, the overall ANOVA on \( \langle g \rangle \) showed a very large effect size. Thus, if students are going to abandon misconceptions, it appears instructors must specifically tell them that preconceived notions are incorrect and then immediately provide clear evidence demonstrating the correctness of the new information.

In addressing the value of refutational text independently of refutational lecture, the conclusions are not as clear. Coverage by refutational lecture together with refutational text was superior to no lecture together with refutational text, which was equal to coverage with standard lecture and standard text. That is, refutational text alone produced about as much change as standard text and standard lecture together. This finding suggests that despite the value of refutational text, the refutational lecture alone produces greater change compared to text alone. Although this finding might suggest that students are not helped as much in reducing misconceptions by reading refutational text as they are by hearing a refutation lecture, the design of the study in the natural classroom environment does not eliminate alternative explanations.

In this context, we assigned readings to students as part of their introductory psychology course; we do not know the degree to which students actually completed the assigned readings. We emphasized the need to keep up with the readings and provided incentives to do so. Nevertheless, the nature of our comparison conditions leaves an assessment of this question unclear. We are currently conducting experimental studies, similar to those conducted by reading education researchers (Marshall, 1989). Such laboratory studies allow greater control over what students read and allow researchers to more clearly assess the value of refutational text. We are also looking at whether students who differ in ability differ in the degree to which they benefit from refutational pedagogy.

In addition, this study has other limits specific to the natural classroom environment. Because this study took place within a normal introductory psychology course, it was important that we cover the usual course content. As a result, we could not use a completely random method to determine the coverage method. We targeted some items (e.g., the 10% myth) because they represent major misconceptions in psychology. We did not target some items (e.g., baby sign language) because they were not essential to understanding the basic principles of psychology. Finally, we were simply unable to find existing readings of the proper type (refutational or standard) for some of the topics. For example, with regard to DARE, we found only refutational text. The experimental studies we have started address this limitation by randomly assigning students to read refutational or standard text we created specifically for the study.
Despite limitations, we believe this study contributes to the literature on psychological misconceptions. Much of the past research on misconceptions has focused on measuring or dispelling individual misconceptions. Although activities are effective in reducing individual misconceptions, it is impractical to think that instructors can address all misconceptions with classroom activities. By relating research on psychological misconceptions to the larger literature on science education and reading research, this study suggests a more general approach to dispelling misconceptions. Targeting misconceptions with a refutational approach, a method shown to be valuable in the education literature, also appears to be a valuable method for dispelling psychological misconceptions. Identifying the factors that influence meaningful change is critical if students are to relinquish popular notions of human behavior and develop a more scientific understanding of psychology.

References


1. We thank Ken Keith for his helpful comments and suggestions.
2. Portions of this article were presented at the 87th Annual Meeting of the Pacific Division of the American Association for the Advancement of Science, San Diego, June 2006.
3. Send correspondence to Patricia Kowalski, Department of Psychology, University of San Diego, 5998 Alcalá Park, San Diego, CA 92110; e-mail: kowalski@sandiego.edu
Do Student Perceptions of Diversity Emphasis Relate to Perceived Learning of Psychology?

Joelle D. Elicker, Andrea F. Snell, and Alison L. O’Malley

The University of Akron

We examined the extent to which students’ perceived inclusion of diversity issues in the Introduction to Psychology course related to perceptions of learning. Based on the responses of 625 students, multilevel linear modeling analyses revealed that student perceptions of diversity emphasis in the class were positively related to how well students believed they understood concepts and the extent to which they believed they learned concepts they could apply to their lives. We also examined the relation between individual differences (e.g., age, race) and perceived learning. We discuss the importance of including issues of diversity in psychology classes.

Psychology educators are increasingly aware of the importance of incorporating multiculturalism into the classroom (e.g., American Psychological Association, 2003; Iijima Hall, 1997; Simoni, Sexton-Radek, Yescavage, Richard, & Lundquist, 1999). Ocampo et al. (2003) urged researchers to conduct empirical work examining the incorporation of issues of diversity in teaching. We sought to answer this call by examining the extent to which students’ perceived learning related to their perceived inclusion of issues of diversity in Introduction to Psychology classes.

The inclusion of multiculturalism in the classroom may facilitate learning in several ways. First, it may increase the extent to which diverse students are able to see the personal relevance of concepts (Rogers, Kuiper, & Kirker, 1977). Second, providing multiple examples of psychological concepts facilitates the development of richer, more accurate concept networks (Siegler & Ellis, 1996). Third, presentation of multiple and diverse examples fosters critical thinking as students reconcile how psychological phenomena operate in different cultural contexts (Carmichael & Hayes, 2001). Examination of multicultural issues is relevant to numerous topics in every major theme covered in the introductory course, as Trimble, Stevenson, and Worrell’s (2003) American Psychological Association commission report documents.

People must integrate new experiences with pre-existing conceptualizations (i.e., schemas) to learn (Carmichael & Hayes, 2001). Integration requires activation of the prior conceptualizations, which vary for students depending on their cultural experiences (Matsumoto, 2000). Hence, discussing the ways in which psychological phenomena occur in multicultural contexts should increase the likelihood that diverse students will see the personal relevance of a concept and activate the necessary schemas. Furthermore, providing multiple examples enables students to observe covariation between features of the examples (Carmichael & Hayes, 2001) and thereby more accurately identify defining elements of concepts.

As a first step in examining the relation between diversity emphasis and learning, our investigation assessed students’ perceptions of these constructs. Students’ perceptions are important because if students believe they can better apply concepts to their lives (i.e., see the personal relevance) when diversity is emphasized in the course material, they will pay more attention to the material and engage in more elaborate and organized processing, which facilitates learning (Klein & Loftus, 1988). Specifically, we proposed that student perceptions of diversity emphasis would be positively related to (a) perceptions of how well they understand concepts (Hypothesis 1) and (b) perceptions that they learned concepts they can apply to their lives (Hypothesis 2), independent of demographic characteristics.
Method

Participants

Undergraduate students (N = 625; 40% men, 59% women, 1% not reporting gender) recruited from Introduction to Psychology classes at a Midwestern university participated in this study. Participants ranged in age from 18 to 64 (M = 20.80, SD = 5.80). Eighty percent identified as European American, 10.6% as African American, 2.9% as biracial or mixed, 2.4% as Asian American, 1.3% as other, 1% as international students and less than 1% as Hispanic or Native American. Three percent self-identified as lower class, 13% as lower middle class, 59% as middle class, 23% as upper middle class, and 1% as upper class. Most students (96%) self-reported as heterosexual or mostly heterosexual. Ninety-nine percent of students indicated they expected to pass the class.

Measures

We found no preexisting measure for perception of diversity emphasis within a course. To assess this variable, we designed a five-item measure (see Table 1). Given the high internal consistency (α = .88) and a principal components factor analysis indicating a one-factor structure explaining 59% of the variance, we created a composite score by averaging the items. We used two single-item measures for the perceived learning variable (see Table 1). Participants provided their age, gender, race or ethnicity, social class standing, sexual orientation, number of absences, and expected grade. Due to the small number of students reporting as members of minority groups, we collapsed race and ethnicity into two categories: European American and non-European American.

Procedure

Students completed surveys anonymously during Week 13 of a 15-week semester. Each section received a code number for tracking purposes. The course is standardized in terms of the readings, assignments, and tests; however, students take the course in small sections of approximately 40 students with different instructors. The instructors’ standardized materials include lecture notes containing multicultural examples for each module as well as discussion questions and activities intended to help students think about the influence of culture on the psychological experience. All sections of the course met two or three times per week, except one section that met on Saturdays.

Data Analysis Strategy

Typically, we would test our hypotheses with ordinary least squares regression. However, a key assumption of regression is that observations (i.e., students) are independent (i.e., randomly sampled from a known population). However, as is typical with classroom research, students may be clustered on the outcome variables largely because these variables are related to demographic characteristics such as race and age, which are not randomly dispersed across classes (e.g., working students usually take classes in early mornings or evenings). As such, regression analysis is inappropriate for our research questions. Fortunately, multilevel linear modeling (MLM) can remove class group effects, as well as other confounding variables, before testing the effects of our variable of interest: students’ perception of diversity emphasis in the course.1

---

Table 1. Study Variables and Descriptive Statistics

<table>
<thead>
<tr>
<th>Study Variables</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception of diversity emphasis in class (scale alpha = .88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How well do you feel this course addressed issues of diversity?</td>
<td>3.27</td>
<td>.88</td>
</tr>
<tr>
<td>To what extent were examples from different cultures given in lectures?</td>
<td>3.33</td>
<td>.88</td>
</tr>
<tr>
<td>To what extent were you asked to think about experiences in different cultures?</td>
<td>3.10</td>
<td>.88</td>
</tr>
<tr>
<td>To what extent did your instructor present questions that made you consider people from other backgrounds?</td>
<td>3.38</td>
<td>.89</td>
</tr>
<tr>
<td>To what extent were questions about different cultures included in class assignments and tests?</td>
<td>2.94</td>
<td>.89</td>
</tr>
<tr>
<td>Perceived learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>To what extent do you feel you understood the concepts presented in this course?</td>
<td>3.83</td>
<td>.79</td>
</tr>
<tr>
<td>To what extent did you learn things in this course that you can apply to your life?</td>
<td>3.70</td>
<td>.88</td>
</tr>
</tbody>
</table>

Note. N = 625. All items were rated on a scale ranging from 1 (not at all) to 5 (a great deal).

---

1See Tabachnick and Fidell (2007) for a more comprehensive comparison of traditional regression and analysis of variance approaches to MLM.
Table 2. Multilevel Linear Modeling Results

<table>
<thead>
<tr>
<th>Perceived Learning Measures</th>
<th>Understand Concepts</th>
<th>Apply to Life</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1: Class effect?</strong></td>
<td>( \varphi_{oo} )</td>
<td>( \varphi_{oo} )</td>
</tr>
<tr>
<td>( \varphi_{oo} )</td>
<td>2.37*</td>
<td>.04</td>
</tr>
</tbody>
</table>

| **Step 2a: Single Control Variable Analyses—Does a potential control variable relate to each DV?** | \( t \) | \( \varphi_{oo} \) | \( t \) |
| Race/ethnicity* | -.28 | -3.56*** | -.01 | -1.118 |
| Sex\(^b\) | .03 | .41 | -.07 | -.122 |
| Age | .02 | 3.02** | .02 | 3.50** |
| Socioeconomic status\(^c\) | .03 | .70 | .03 | .72 |
| Sexual orientation\(^d\) | .11 | 2.28* | .01 | .25 |
| Expected grade\(^e\) | .47 | -13.85*** | .23 | -5.53*** |
| Absences | -.06 | -2.96** | -.04 | -1.54 |

| **Step 2b: Overall Control Model—Which control variables significantly predict in combination?** | \( t \) | \( \varphi_{oo} \) | \( t \) |
| Age | .01 | 2.63** | -.02 | 3.23** |
| Expected grade | .46 | -13.37*** | .22 | -5.22*** |
| Sexual orientation | .13 | 2.73** | — | — |

| **Step 3: Final Model—Does perception of diversity emphasis predict beyond the control model?** | \( \varphi_{oo} \) | \( \varphi_{oo} \) |
| Age | .01 | 2.43* | .02 | 3.07*** |
| Expected grade | .43 | -12.91*** | .18 | -4.43*** |
| Sexual orientation | .13 | 2.95** | — | — |
| Perception of diversity emphasis | .26 | 6.95*** | .42 | 9.28*** |

Note. \( N = 625 \).
\(^*\) European American, \(^1\) non-European American. \(^*\) male \(^1\) female. \(^\circ\) lower class, \(^5\) upper class. 
\(^d\) exclusively homosexual, \(^5\) exclusively heterosexual. \(^*\) \( F \), \(^5\) \( A \).

Results

MLM is a model building technique that requires multiple analyses rather than a single analysis. First, we removed the nested data effect of students embedded in particular classes (Step 1 in Table 2 shows the significant class effects). Next, we built a model of unique predictive control variables (Step 2) and then determined if perceptions of diversity emphasis predicted above and beyond these variables and class effects (Step 3).

We built the overall control model in two phases. First, we assessed potential control variables in response to Ocampo et al.’s (2003) suggestion that students from diverse backgrounds might differ in their experiences within a classroom setting. Additionally, we anticipated that number of absences and students’ expected grade would relate to perceived learning. Thus, we independently entered each potential control variable as a predictor of each of our perceived learning measures (Step 2a in Table 2). Second, we created the overall control model by determining which of these significant control variables in combination significantly predicted the perceived learning variables, using a stepwise-type of analysis by starting with the most predictive variables (from Step 2a) and adding additional variables. For the student understanding of concepts variable, the overall control model included age, expected grade, and sexual orientation. For the student application to life variable, age and expected grade were the only uniquely predictive variables (see Step 2b of Table 2).

The overall control model provided the necessary covariates to our hypothesis tests, which we conducted by adding perception of diversity emphasis in the classroom to see if this variable predicted each perceived learning measure above and beyond the necessary control variables and the class differences. As shown in Step 3 of Table 2, the Gamma coefficient was significant for both the student understanding of concepts and student application to life variables. The Gamma coefficient is analogous to the unstandardized b weight in a regression analysis and reflects the presence of
unique prediction. Thus, the results of Step 3 supported both hypotheses that student perceptions of diversity significantly predict understanding and relevance. To examine the size of these significant effects, we calculated the amount of variance that perception of diversity explained for individuals who were in the same classes (Kreft & de Leeuw, 1998). Students' perceptions of diversity explained 6% of the variance in the within-class differences for understanding concepts presented in class. Students' diversity perceptions had a greater impact on their perceived relevance of the course, as 11% of the variance was explained. These effects reflect explained variance of students' perceptions of diversity above and beyond the control variables in the control model. Thus, students' perceptions of diversity had almost twice the effect on how they saw the course content being applicable to their lives in comparison to how well they understood the concepts of the course.

**Discussion**

Consistent with our hypotheses, student perceptions of the extent to which the course emphasized issues of diversity were positively related to reported understanding of course concepts and reported ability to apply course content to life experiences. These relations held even when controlling for demographic (e.g., age, race, sex) and pertinent contextual characteristics (e.g., course section and expected grade). Thus, regardless of students' race, ethnicity, or other characteristics, if students perceived more diversity in the curriculum, they believed that they understood psychological concepts more and could better apply concepts to their lives. Students' perception of diversity emphasis was the largest contributor to their belief that course material was applicable to their lives. Not surprisingly, for understanding of course concepts, students' expected grade was the most important predictor, followed by diversity perceptions. The statistically significant effects of student perceptions of diversity on their understanding of the course material and course relevance support our contention that diversity is an important part of the classroom experience. We do not want to overstate the importance of this variable, however, because the effect sizes are relatively small (Cohen, 1988). These data do point to the need for instructors to infuse multicultural examples into their courses to increase the personal relevance of course content to students' lives. Seeing the personal relevance of material enables self-referent encoding, which yields superior memory due to elaborate processing (Klein & Loftus, 1988; Rogers et al., 1977; Symons & Johnson, 1997).

Our findings represent a first step in disentangling the role of individual differences and class-level differences in investigations of diversity emphasis and perceived learning. Our conclusions should be considered with caution, as they are based on self-report data; future research should assess objective measures of learning and diversity emphasis. Additionally, research with more heterogeneous samples is needed to better understand the relation of being a member of underrepresented groups to perceived diversity emphasis. Finally, the correlational nature of our results warrants cautious interpretation and points to the need for experimental research to clarify potential causal interpretations. Thus, our findings should be taken as preliminary rather than conclusive.

It is interesting that two underrepresented groups, non-European Americans and students identifying as homosexual, reported less understanding of course concepts than those in the majority. Our data suggest that students from diverse backgrounds might indeed have different experiences within the classroom, thereby supporting Ocampo et al.'s (2003) call to examine the learning experiences of students from diverse backgrounds. We recognize that diversity in a course curriculum is but one influence on students' perceptions of diversity emphasis; structural diversity (i.e., racial and ethnic composition of a class) and informal interactions transpiring outside the classroom play important roles as well (Hurtado, Dey, Gurin, & Gurin, 2003). Nevertheless, the groups' different experiences point to the need for interventions that incorporate diversity in ways specifically targeted to support these groups and eliminate any disparity in perceived or actual learning. As educators, we know of the ethical call to incorporate multiculturalism into the classroom (American Psychological Association, 2002). By linking perceived diversity emphasis to perceived learning, our findings provide yet another reason to pay attention to issues of diversity.

**References**


American Psychological Association. (2003). Guidelines on multicultural education, training, research, practice, and

Notes

1. We thank Mindi Thompson for her help with measure development and data collection.
2. Send correspondence to Joelle D. Elicker, Department of Psychology, The University of Akron, Akron, OH 44325–4301; e-mail: joelle@uakron.edu.
Does the First Week of Class Matter? 
A Quasi-Experimental Investigation of Student Satisfaction

Anthony D. Hermann
Bradley University

David A. Foster
Western Oregon University

Erin E. Hardin
Texas Tech University

Teaching experts suggest that establishing clear expectations and a supportive environment at the beginning of a college course has a lasting impact on student attitudes. However, minimal empirical evidence exists to support these suggestions. Consequently, we randomly assigned instructors to either begin their course with a reciprocal interview activity aimed at these goals or in their typical fashion. At term’s end, students experiencing the activity \((n = 187)\) reported greater clarity regarding their course responsibilities, more support from their instructor, and greater course satisfaction on both official evaluations and experimenter-administered measures, compared to students who had not \((n = 190)\). These results contribute to a converging body of evidence regarding the effectiveness of reciprocal interviews and similar activities generally.

Teaching experts frequently assert that the first days of a college course have a long-lasting impact on the classroom environment and student attitudes. Wilson and Wilson (2007) were the first to empirically examine the effect of the first day of class on students’ subsequent course evaluations and grades. Students who experienced a positive first day (i.e., a 15-min video of a friendly instructor who dismissed class early) reported more positive perceptions of the professor and more motivation for the course compared to those students who experienced a negative first day (i.e., a boring videotaped instructor who used all class time and assigned homework). The motivation differences persisted, and positive condition students had higher grades at term’s end.

This innovative study suggests that the long-held belief in the importance of the first day of class is appropriate. More specific guidance is needed, however, regarding how instructors can effectively establish a positive and productive environment. It remains unclear, for example, whether Wilson and Wilson’s (2007) findings resulted from the differences in instructor warmth, homework assigned, or how class time was utilized. They crafted their positive and negative sessions based on students’ stated preferences, but there is little evidence connecting these preferences to student satisfaction and other outcomes. Moreover, there is great variability in the content and form of first-day activities. Although there is some evidence that students dislike icebreakers (Henslee, Burgess, & Buskist, 2006), some ice-breaking activities might be more effective than others. Thus, it is important
to investigate what specific activities create favorable environments.

Common expert suggestions for establishing a positive and productive learning environment include both making instructor course expectations clear (Curzan & Damour, 2000; Davis, 1993) and creating a dynamic and supportive classroom community (Lucas, 2006; McKeachie & Svinicki, 2006; Royse, 2001). To accomplish these goals, Hermann and Foster (2008) proposed a reciprocal student–instructor interview activity, adapted from organizational psychology textbooks (Harvey & Brown, 2000; Osland, Kolb, & Rubin, 2000), in which the instructor solicits information from the students and the students then collectively ask questions of the instructor. In addition to clarifying course expectations, the interview aims to make students more comfortable interacting with the instructor and each other.

Two studies have explored the immediate impact of reciprocal student–instructor interviews (Case et al., 2008; Hermann & Foster, 2008). Among the findings, students reported that (a) they enjoyed the activity; (b) the activity clarified the instructor’s expectations; and (c) they felt more comfortable participating in class and interacting with the instructor. To date, however, no research has demonstrated the long-term impact of this or any other first-week activity. Given the favorable responses observed in previous studies, we predicted that students who had experienced the activity would be more satisfied with the course at the end of the term than students who had not. Second, given the activity’s main goals, we hypothesized that student perceptions of instructor support and expectation clarity should also be greater for those experiencing the activity and that these variables should account for a significant portion of the activity’s effect on satisfaction. To answer these questions, we used a quasi-experimental design and randomly assigned 16 sections of introductory psychology students to experience the activity or not and then assessed outcomes at the end of the 15-week term.

Method

Participants

Participants were 377 undergraduate students (age $M = 19.8$ years, $SD = 3.8$ years; 56% female) enrolled in 1 of 16 sections of introductory psychology at a large Southwestern university. Ten graduate teaching assistants taught 16 sections. Seven of the sections met twice weekly for 80 min over a 15-week term, whereas the remaining met three times weekly for 50 min. Instructors (and hence their students) were randomly assigned to an activity ($n = 187$) or no activity condition ($n = 190$). The instructors were relatively inexperienced ($M = 1.6$ semesters prior teaching, mode = 2), and prior experience did not differ between groups, $p > .40$.

Measures

Clarity and supportiveness. These measures assessed the degree to which students had come to expect particular behaviors from their instructor at the end of the term. Students rated the degree to which the instructor communicated specifically and unambiguously regarding course expectations on five clarity items (e.g., “I expect from my instructor that he or she . . . specifically describes the evaluation criteria in this course”). Likewise, they rated the degree to which the instructor supported and appreciated student effort on five supportiveness items (e.g., “treats me as a person, not a number”). All items were assessed using a 5-point response range from −2 (entirely disagree) to 2 (entirely agree). Both scales yielded adequate internal reliabilities ($\alpha = .84$ and .88, respectively).

Satisfaction with course. Participants also indicated their “overall satisfaction with this course” on a single item using a 7-point scale of −3 (very dissatisfied) to 3 (very satisfied). We also collected mean ratings for all 16 items on the official university student evaluations of instruction for each section, which were rated on a 5-point scale of 1 (strongly disagree) to 5 (strongly agree). We were particularly interested in items about course satisfaction and the activity’s goals of clear expectation and establishing dialogue (see Table 1).

Procedure

Instructor training. During the week before classes, each group of instructors participated in separate orientation sessions. Both groups were instructed to conduct a typical first day (i.e., syllabus overview, brief icebreaker, brief introduction to course material) and the experimental condition was also given instruction on conducting the reciprocal interview activity. To keep instructors blind to the study’s purpose, all were asked not to speak with other instructors about
Table 1. Effects of Experimental Condition on University- and Experimenter-Administered Measures

<table>
<thead>
<tr>
<th>Items and Source</th>
<th>Activity Condition</th>
<th>No Activity Condition</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>University items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall this course was a valuable learning experience.</td>
<td>4.55</td>
<td>0.27</td>
<td>4.12</td>
</tr>
<tr>
<td>The instructor welcomed and encouraged questions and comments.</td>
<td>4.81</td>
<td>0.10</td>
<td>4.50</td>
</tr>
<tr>
<td>Expectations were clearly stated either verbally or in the syllabus.</td>
<td>4.76</td>
<td>0.10</td>
<td>4.38</td>
</tr>
<tr>
<td>Experimenter items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectation clarity</td>
<td>1.21</td>
<td>0.62</td>
<td>1.02</td>
</tr>
<tr>
<td>Supportiveness</td>
<td>1.14</td>
<td>0.64</td>
<td>0.84</td>
</tr>
<tr>
<td>Satisfaction with course</td>
<td>2.31</td>
<td>1.04</td>
<td>1.94</td>
</tr>
</tbody>
</table>

Note. Sample sizes for university items were 8 in each condition and 187 and 190 for the experimenter items (activity and no activity conditions, respectively).

the study until it was completed.\(^1\) All instructors were informed of the importance of student perceptions and expectations and that these perceptions would be measured. All sections used the same textbook, were required to cover certain content, and gave similar assignments.

**Activity.** Students participating in the reciprocal interview activity formed small groups of 5 or 6 and had approximately 10 to 15 min to discuss several course-related issues. The instructors explained that the discussion was preparation for an interview, and each group selected a representative to field the instructors' questions and represent their groups' responses. Guided by a handout, the groups discussed a range of topics including expectations, goals, and experiences related to the course; suggestions for classroom norms; and instructor behaviors that could help them achieve their goals. Immediately afterward, the instructors interviewed the group representatives in the presence of the class. Instructors conveyed interest by taking notes (on blackboard or notebook) and by asking clarifying questions.

Immediately after the instructor interview, the groups were asked to elect new representatives to interview the instructor on the group's behalf. The groups were given 5 to 10 min to agree on several questions, guided by topics on the handout (e.g., the instructor's expectations, evaluation practices). Students were encouraged to ask any question related to the course. When responding, instructors answered thoughtfully and sincerely and promised to return to issues if they needed additional time. This also provided opportunities to cover important issues that had not yet been addressed, like the challenging course aspects or the ways to get assistance with course material.

During the last week of the course, research assistants administered the dependent measures to all students. After the term ended, the researchers collected official course evaluation summaries from each section.

**Results and Discussion**

We assessed differences between the experimental and control groups using MANOVA on the measures collected at term's end. First, we analyzed mean ratings from each section (N = 16) to examine differences on the university evaluation items. Second, we examined mean differences on the ratings of expectations of clarity, supportiveness, and satisfaction that we administered to students individually.

**University-Administered Measures**

As Table 1 displays, on the university evaluations, sections that experienced the activity reported more favorable attitudes about the course. For example, activity sections rated the course as a more valuable learning experience (M = 4.55, SD = .27) than the nonactivity sections (M = 4.11, SD = .36), F(1, 14) = 8.59, p = .01, d = 1.20. Likewise, activity sections perceived that

\(^1\)Although instructors in the no activity condition were not given specific instructions to avoid other introductory activities to avoid alerting suspicion to the study's purpose, there is little reason to believe these instructors engaged in a similar activity on their own. The course supervisor (who is also the third author) reviewed all syllabi to ensure uniformity in content and requirements and found no evidence that any instructor in the no activity condition devoted a significant amount of time to any type of extensive expectation-clarifying activity.
their instructor welcomed questions more (M = 4.81, SD = .10) than the nonactivity sections (M = 4.50, SD = .33), F (1, 14) = 6.43, p = .02, d = 1.09, and that expectations were more clearly communicated (M = 4.76, SD = .10) than in the nonactivity sections (M = 4.38, SD = .23), F (1, 14) = 17.69, p = .001, d = 1.45. Moreover, the same effects were observed on 12 of the remaining 13 university-administered items (ds = .99–1.50; all ps < .05) with the only exception regarding ratings of workload (a constant across all sections; d = .83, p = .10). This consistent and robust set of effects is particularly noteworthy given that conducting analyses on means (rather than scores) reduces the effects of outliers and using section as the unit of analysis afforded limited statistical power.

**Experiment-Administered Measures**

At the individual level, as expected, students who experienced the activity also reported more satisfaction with the course on the experimenter-administered measures (M = 2.31, SD = 1.04) than those who did not (M = 1.94, SD = 1.32), F (1, 369) = 12.77, p < .01, d = .31. Additionally, as expected, students experiencing the activity had come to expect more clarity (M = 1.21, SD = .62) than those who did not (M = 1.02, SD = .69), F (1, 369) = 7.73, p < .01, d = .29) and more supportiveness (M = 1.14, SD = .64) than those who did not (M = .85, SD = .78), F (1, 369) = 15.82, p < .001, d = .42.2 These differences were not moderated by student gender (ps > .20). Ratings of supportiveness were strongly and positively correlated with ratings of expectation clarity (r = .59, p < .001), whereas each of these variables was moderately correlated with course satisfaction (supportiveness r = .38, p < .001; clarity r = .31, p < .001).

Next, using linear regression, we performed mediational analyses on the experimenter-administered items to determine whether the activity’s effect on student satisfaction could be explained by either clarity or supportiveness. Without mediators entered, activity condition was positively related to course satisfaction (β = .15, p < .01) and accounted for 2.4% of the variance. Expectation clarity partially mediated this relationship (partial condition β = .11) with condition accounting for a significantly lower portion of the variance regarding course satisfaction with clarity added to the equation (1.2%; Sobel’s test z = 3.57, p < .001; see Figure 1). Similarly, supportiveness also partially mediated the relationship between the activity condition and satisfaction with the course (partial condition β = .08) with condition accounting for significantly less portion of the variance regarding course satisfaction with supportiveness added (0.6%; Sobel’s test z = 2.57, p < .05; see Figure 2). Finally, stepwise regression analyses showed that both clarity (∆R² = .01, p < .05) and supportiveness (∆R² = .06, p < .001) explained unique variance in satisfaction with course, even when controlling for the effects of the other (total R² = .16). These findings suggest that, by using a fairly simple intervention in the first week, instructors can create a positive environment that has long-lasting effects on student perceptions of the instructor and course satisfaction. Students who experienced the reciprocal interview expected a more supportive learning environment and more clear communication from their instructors at the end of the term, which accounted in part for their higher level of satisfaction. The findings reported here and those of previous studies demonstrating the positive proximal effects of this activity (Case et al., 2008; Hermann & Foster, 2008) provide converging support of teaching experts’ recommendations

---

2 Analyses of these variables using course section as the unit of analysis yielded a very similar pattern of findings with activity sections reporting more satisfaction (M = 2.36, SD = .31 vs. M = 1.96, SD = .61, d = .78), more clarity (M = 1.25, SD = .16 vs. M = 1.03, SD = .13, d = 1.21), and more supportiveness (M = 1.16, SD = .19 vs. M = .88, SD = .24, d = 1.09) than nonactivity sections, F (1, 14) = 2.67, p = .12; F (1, 14) = 9.08, p = .01; and F (1, 14) = 6.45, p = .02, respectively.
regarding the importance of the first week of class.

Although it is clear that this activity has a demonstrable impact on student satisfaction, it remains unclear whether the activity also impacts student performance. Likewise, although it seems plausible that clearer communication of expectations and a more supportive environment could translate into better performance, these factors could also lead to higher expectations on the part of the instructor or happier but less productive students. It also remains unclear which aspect of the activity is responsible for the observed effects. It could be the instructor’s attention or the reciprocal exchange that is the key ingredient. Moreover, it might be that students’ perceptions of the instructor’s intentions for the activity matter most. Future research can and should provide more evidence about which aspect of first-week activities has the most impact. Likewise, future research can elucidate which types of courses benefit from which types of activities. This study focused on a lecture-oriented introductory social science course targeted at first- and second-year students, but it remains unclear if the activity would have more or less of an impact in smaller, upper level courses or courses where interaction might be less common (e.g., history and systems, physiological psychology).

The reciprocal interview activity explored here addresses a number of student issues that might also contribute to a more satisfying experience. Giving students an early opportunity to meet classmates, requiring early and active participation, and normalizing concerns through public discussion might all affect their experiences (see Henslee et al., 2006). Regardless of how instructors strive toward these objectives, more research is needed to gain an empirical understanding of the mechanisms by which the first days of class impact long-term course outcomes.

Another intriguing issue raised by the reciprocal nature of this activity is the degree to which it might affect instructor behavior and motivation. Like their students, instructors who use the activity might better understand others’ perspectives (e.g., concerns, valued behaviors, goals). As a key component of caring for others, such understanding could have measurable effects on student outcomes and evaluations (Teven & McCroskey, 1996). Moreover, such caring has shown to be related positively to instructor motivation (Teven, 2007). Although little empirical work links instructor motivation to student outcomes, one study of secondary students demonstrated that teacher motivation positively predicted future student achievement, even controlling for baseline achievement (Knowles, 1999). Thus, the reciprocal interview might not only provide information helpful in creating an effective learning environment, but also induce a stronger commitment to do so. Anecdotal evidence from this study’s instructors and our own more seasoned experience suggests that the activity has an energizing effect. The activity requires the instructor to think deeply about his or her objectives in preparation and then creates a direct and lively exchange about those objectives with a receptive student audience.

Although this study used only relatively new instructors (and the current findings might only apply to them), the activity can be useful for different reasons depending on the instructor’s level of experience and how many times he or she has taught the course. For relatively new courses or instructors, it might accelerate the learning curve about student perspectives on the course. For instructors with more experience, it could be a way to gain fresh perspective on and motivation for a well-worn course. We look forward to future research that will shed light on the impact of the first week of class on student satisfaction and achievement as well as innovative ways to promote clear communication and a supportive environment in the college classroom.

References


Notes

1. The authors would like to thank Tanecia Blue, Timothy Ballew, Erin Buck, Dani Citti, Kelly Davis, Tammy Ott, Amy Pietan, Shawn Rose, Rocio Villareal, and Zach Ward for allowing us access to their courses to conduct the study and to Carol Jackson and Femina Varghese for their help with data collection.

2. Send correspondence and requests for additional details or clarification about the activity to Anthony D. Herrmann, Department of Psychology, Bradley University, 1501 W. Bradley Avenue, Peoria IL 61625; e-mail: ahermann@bradley.edu.
FACULTY FORUM

Exploring Interdisciplinary Themes in Introductory Psychology

Kristin A. Ritchey
Department of Psychological Science Ball State University

Jennifer P. Bott
Department of Marketing and Management Ball State University

We describe a method for helping introductory psychology students identify interdisciplinary connections among 5 social science disciplines. Pre- and posttest data assessed 359 undergraduates’ understanding of psychology’s relation to other fields. Results indicate the method is effective and provides one way for individual instructors to address interdisciplinary themes in their courses.

For many years instructors have been concerned that students leave college without understanding the interconnections of their courses (Parisi, 1985). Previous approaches to teaching interdisciplinary connections have involved large-scale curriculum changes (e.g., Allen-Meares, 1998; Golding & Kramer, 2000), team-teaching (e.g., Widner & Davies, 2007), campus-wide learning communities, (e.g., Rashne, 2009), and first-year seminars (e.g., Goodman & Pascarella, 2006; Porter & Swing, 2006). However, as Brownell and Swaner (2009) noted, despite the ample research showing that first-year seminars and other experiences focusing on course integration are beneficial, the reality is that many students do not have those experiences. We suggest a method of implementing interdisciplinary materials in an introductory psychology class that can be used by individual instructors without large scale revisions.

Current Research

Our goal was to increase introductory psychology students’ ability to recognize and provide examples of how psychology influences and is influenced by other disciplines. To this end, the first author adapted Weiten’s (2005) theme “Psychology evolves in a socio-historical context” into the question “How does psychology SHAPE your world?” with the SHAPE acronym standing for sociology, history, anthropology, political science, and economics. We implemented this question as an explicit theme of the course. Our purposes for altering the wording of Weiten’s theme were to give students concrete examples of five disciplines that contribute to the “socio-historical context”; to emphasize that not only does psychology develop within this context, but also actively influences this context, or shapes the world in which we live; and to encourage students to think of this theme as a question that needed to be investigated as opposed to a statement that required no answer. We hypothesized that students who completed this section of introductory psychology would identify more examples of how psychology related to other disciplines than a control group of students completing sections of introductory psychology that did not address the theme; that these
students would rate psychology’s influence on other fields greater than the control group; and that they would rate other fields’ influence on psychology greater than the control group.

Method

Participants

Students (226 women, 133 men) enrolled in a large, Midwestern university participated. Of the 359 participants, 201 were enrolled in the section taught by the first author and the remaining 158 were enrolled in roughly equal numbers (41 in Section 1, 38 in Section 2, 44 in Section 3, and 35 in Section 4) in sections taught by other instructors. These four sections served as control groups. Participants self-identified as 44% Caucasian, 33% Hispanic, 18% Black, 2% multiracial, 0.5% Asian, 0.5% Native American, and 2% other. Ages ranged from 18 to 26 years (M = 19 years, SD = 1.25 years). The distribution of class standing was 58% first-years, 27% sophomores, 10% juniors, and 5% seniors. Participants’ majors represented every college within the university, with the mode (24%) being undecided. Students indicated that 37% had been enrolled in a previous psychology course. Students in the experimental group received extra credit toward their final course grade for their participation. Students in the control group received credit toward fulfillment of a course requirement.

Materials

Course materials. A variety of materials were utilized in the course, including videos, newspaper articles, advertisements, photographs, quotations, research articles, children’s books, and book excerpts. Materials were gathered by consulting faculty who taught courses in sociology, history, anthropology, political science, or economics, by reviewing recommended readings in introductory psychology texts, and from our own naturally occurring exposure to these materials. Table 1 lists a sample of these materials, the discipline to which they were related, and the connection made between the materials and course content.

Survey materials. We created a 14 item pre- and posttest instrument. The “demographic” section included six items measuring age, sex, race, major, class standing, and whether students had taken a previous psychology course. The “interdisciplinary” section included six short-answer questions that asked students to provide an example of a topic or concept from psychology that might also be discussed in sociology, history, anthropology, political science, economics, and one other class of their choice. Two Likert scale questions measured students’ perception of the extent to which psychology influences other fields or is influenced by other fields using a 1 (no influence) to 5 (very strong influence) response range.

Procedure

During the first class meeting of a semester students in the experimental group were read a script requesting that they complete the survey but were told they were free not to do so. During the rest of the class meetings of the semester, the first author posed the question “How does psychology SHAPE your world?” After posing the question I provided explicit examples of psychology’s influence on one of the SHAPE disciplines or an example of how psychology was influenced by one of those disciplines. This influence was demonstrated on the PowerPoint slides and by a brief (approximately 5–10 min) class discussion of this topic in which I would provide further clarification, students would ask questions, and students would contribute comments. Between two and five examples of each of the SHAPE disciplines were used throughout the semester. This variability was due in part to how much overlap existed between the other disciplines and the course content and due in part to students providing their own examples of interdisciplinary materials. (This was not required, but provides anecdotal evidence of students’ engagement with the course and the method’s effectiveness). We collected the posttest data on the last day of the semester, using the same survey as the pretest.

We collected the control group data 24 months after the experimental group. The same posttest was made available to introductory psychology students through an online survey system. The online system allowed us to collect data from more students than a paper administration of the posttest, which would have been limited due to lack of class time; the online system was completed at students’ convenience outside of class. Students were first shown an informed consent form, and if they agreed to participate, were directed to the survey. For both paper and online versions, students spent approximately 10 min responding.
Results

Six open-ended survey items asked students to provide examples of how psychology pertained to the five disciplines in the SHAPE acronym and one other discipline of their choice. Two raters, blind to experimental condition, coded responses as 0 (no or incorrect response) or 1 (correct response). The coding rubric was developed by consulting introductory-level textbooks in the five disciplines and professors who taught those courses, resulting in a list of topics that were considered appropriate or likely to be discussed both in psychology and in at least one of those disciplines. When a student's response did not appear on the list, we again consulted a textbook or a professor to double-check whether that topic fell under the purview of that field. The raters independently scored a random subset of 72 protocols, and interrater reliability was high ($\kappa = .90$).

Changes From Pretest to Posttest

We compared the number of students in the course who earned credit for their responses to the six open-ended questions on the pretest and on the posttest. Chi-square analyses show significant differences, with more students being able to correctly address how psychology related to the five focal courses: sociology, $\chi^2(1, N = 201) = 128.14, p < .001, \phi = .79$; history, $\chi^2(1, N = 201) = 58.61, p < .001, \phi = .54$; anthropology, $\chi^2(1, N = 201) = 140.94, p < .001, \phi = .84$; political science, $\chi^2(1, N = 201) = 72.68, p < .001, \phi = .60$; economics, $\chi^2(1, N = 201) = 144.73, p < .001, \phi = .85$; and one other course of their choice, $\chi^2(1, N = 201) = 105.03, p < .001, \phi = .72$, after the class than before the class. Paired-samples $t$ tests were used to examine pre- and postcourse differences in the perception that psychology influences other fields and is influenced by other fields. Students’ perceptions of psychology’s influence in other fields was significantly higher at posttest, $t(155) = -3.021, p < .001, \eta^2 = .06$; as was their belief that psychology was influenced by other fields, $t(150) = -3.020, p < .001, \eta^2 = .08$.

Comparing Posttests

Posttest responses were compared for the students enrolled in the SHAPE course and the students enrolled in the other sections of introductory psychology. An ANCOVA was utilized, with completion of the SHAPE section of introductory psychology versus control sections as the independent variable and participation in a previous psychology course as a covariate. Previous exposure to psychology did not have a significant influence on participants’ responses, $F(11, 265) = 1.35, p = .199, \eta^2 = .11$

Students in the SHAPE course were more able to provide examples of how psychology related to

Table 1. Examples of Materials Used to Address the Theme “How Does Psychology SHAPE Your World?”

<table>
<thead>
<tr>
<th>Topic</th>
<th>Materials</th>
<th>Connection to Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociology</td>
<td>Political quotes regarding need for social programs</td>
<td>Social programs’ influence on development</td>
</tr>
<tr>
<td>History beginning</td>
<td>World War II film clips</td>
<td>Influence on Milgram and of social psychology</td>
</tr>
<tr>
<td>Anthropology</td>
<td>Donald Brown’s Human Universals</td>
<td>Blank slate philosophy and nature vs. nurture</td>
</tr>
<tr>
<td>Political science</td>
<td>News story regarding parental laws</td>
<td>Nature vs. nurture/childrearing</td>
</tr>
<tr>
<td>Economics</td>
<td>Kahneman’s research regarding decision making</td>
<td>Persuasion and emotion</td>
</tr>
<tr>
<td></td>
<td>Excerpts from Fast Food Nation</td>
<td>Maslow’s Hierarchy of Needs</td>
</tr>
</tbody>
</table>
sociology, $F(2, 275) = 3.28, p < .05, \eta^2 = .62$; history, $F(2, 275) = 4.75, p < .01, \eta^2 = .70$; and political science, $F(2, 275) = 12.62, p < .001, \eta^2 = .86$, than students not enrolled in the course. Students in the SHAPE course were equally as able as students not in the course to provide examples of how psychology related to anthropology, $F(2, 275) = 2.05, p = .13, \eta^2 = .51$, and economics, $F(2, 275) = 1.86, p = .16, \eta^2 = .48$. Students in the SHAPE course provided higher estimates of the extent to which psychology influences other fields, $F(2, 275) = 7.94, p < .001, \eta^2 = .79$, and higher estimates of the extent to which psychology is influenced by other fields, $F(2, 275) = 7.69, p = .001, \eta^2 = .79$, than students not enrolled in the course.

Discussion

This study examined the influence of incorporating various social science materials into an introductory psychology course on students’ ability to identify and provide interdisciplinary connections. Compared to their own pretest responses and compared to posttest responses of students enrolled in other sections of introductory psychology, students who completed the course were better able to provide examples of psychology’s relation to three other disciplines. A content analysis of the course PowerPoint slides and lecture notes shows examples of psychology relating to sociology, history, and political science were provided an average 4.33 times in the semester compared to an average of only 2 examples for anthropology and economics, which likely contributed to the lack of effect for these two disciplines. Students in the course rated psychology’s influence on other fields and the influence of other fields on psychology higher than the control group and higher than they had on the pretest. Possible limitations include differences between the experimental and control groups’ methods of data collection as well as differences in course instructors.

In sum, we suggest this approach as a relatively easy way for individual instructors to implement an interdisciplinary theme into their courses and to assist students in identifying cross-curriculum connections. This approach helps combat the possibility of students leaving their entire undergraduate experience missing valuable connections between disciplines (Parisi, 1985). Given psychology’s applicability to other disciplines and its ability to achieve the goals of a liberal education (McGovern, Furumoto, Halpern, Kimble, & McKeechiae, 1991), and considering the large number of students exposed to general psychology courses, an introductory psychology course seems a logical and useful place to explore interdisciplinary connections. In fact, recent reflections on undergraduate education suggest that creating psychologically literate citizens (i.e., students who can apply psychological principles to other fields and everyday situations) should be a primary focus of the future of the discipline (McGovern et al., 2009). We believe this approach provides a simple and effective means for answering that call.

References

Notes

1. Portions of this article were presented at the annual meeting of the Association for Psychological Science (May 2007) in Washington, DC.

2. We thank Laura Graves and Amber Lupo for their help collecting and coding the data.

3. Send correspondence to Kristin A. Ritchey, 104 North Quad, Department of Psychological Science, Ball State University, Muncie, IN 47306; e-mail: karitchey@bsu.edu
Exploring Interdisciplinary Themes in Introductory Psychology

Kristin A. Ritchey \textsuperscript{a} & Jennifer P. Bott \textsuperscript{b}
\textsuperscript{a} Department of Psychological Science, Ball State University
\textsuperscript{b} Department of Marketing and Management, Ball State University

Available online: 11 Oct 2010

To cite this article: Kristin A. Ritchey & Jennifer P. Bott (2010): Exploring Interdisciplinary Themes in Introductory Psychology, Teaching of Psychology, 37:4, 262-266

To link to this article: http://dx.doi.org/10.1080/00986283.2010.510958

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Across different terms, introductory psychology college instructors offered incentives to students for participating in department-wide research as part of a course requirement. Students received either no incentive (n = 590) or exam “bonus points” (n = 480) on both a midterm and final exam. Students offered bonus points began work at an earlier date and completed the research requirement more frequently than students not offered bonus points. Use of the bonus-point system could benefit students, instructors, and researchers.

Using Exam Bonus Points As Incentive for Research Participation

Joseph R. Ferrari
Stephanie McGowan
DePaul University
Many U.S. psychology departments conduct empirical research using undergraduate participant pools (Rosenthal & Rosnow, 1975; Sieber & Saks, 1989). Participant pools provide researchers with convenient samples for investigating theories and concepts and allow students to act as shareholders in the research process by learning firsthand how psychologists conduct research (Masling, O’Neill, & Jayne, 1981). Usually students in lower division courses (e.g., introductory psychology) become participants through advertisements of studies, choose the studies they wish to complete by signing up, and then receive credit for participation after completing their involvement.

Based on low completion rates of research study requirements for introductory psychology courses at our university, in this study we assessed the effectiveness of an incentive program that rewards students for beginning and finishing the requirement earlier. We believed that this incentive program could be beneficial to instructors, students, and researchers. For instructors, the incentive program could reduce frustration at students’ last-minute attempts to complete the requirement and the workload of grade changes after the course has ended. For students, increased research participation could enhance course material. An incentive program also could benefit researchers by encouraging students to participate earlier in the term.

Participating in research helps students learn about psychology and better understand the research process (Landrum & Chastain, 1995). When required to learn about research, students reported being more satisfied with their experience participating in studies than with writing papers on research reports (Waite & Bowman, 1999). Research participation also increases the value of experiments in students’ minds (Holmes, 1967). Consequently, encouraging students to participate can increase the effectiveness of the introductory psychology course and can assist students as they progress in the psychology major.

Increasing student participation and motivating earlier participation may improve the characteristics of researchers’ samples. Research suggests that participant characteristics and task performance vary with time of participation. Students who ordinarily participate early in the term, compared to students who participate late in the term, tend to be more academically oriented, internally controlled (Evans & Donnerstein, 1974), and intrinsically motivated (Hom, 1987); seek structure in their life (Roman, Moskowitz, Stein, & Eisenberg, 1995); and are likely to be “morning people” (Zelenski, Rusting, & Larsen, 2000). Early-term participants performed better on serial learning and symbolism tasks than later participants (Richter, Wilson, Milner, & Senter, 1981). Given that the time of participation may reflect individual differences among students and that there may be performance differences based on when students take part in studies, encouraging early research participation for all students may minimize individual and performance differences. In addition, early participation could benefit students by providing better access to studies and a broader set of studies from which to choose, which could expose students to a larger breadth of psychological information.

At some institutions, failure to participate results in penalty: automatic course failure, decrease in one’s course grade, penalty to complete additional studies beyond the established number, or an incomplete grade in the course until the student completes the required number of studies. We assessed whether incentives for early compliance were an effective strategy to motivate students, as opposed to typical punishment for noncompliance techniques. We assessed whether giving bonus points on exams would prompt students to meet their course requirement earlier. Other than one study on the positive effect of hypothetical incentives (cash payments) on student participation in stressful research studies (Maughan & Higbee, 1981), we found no published studies on interventions that actually increased student research participation.

Method

Participants, Program, and Procedure

College students (18 to 22 years of age) who were enrolled in a 10-week introductory psychology course offered at a large, urban, Midwestern university participated in the study. The 60-min class met 3 days per week and typically has had high enrollment (100 to 180 students). Students from 12 sections of the course participated in this study. Six different instructors taught the sections.

As part of the course requirement and an established departmental policy, students completed “research hours” to receive a course grade. Although the course instructor determined the number of required research hours, instructors usually require 6 hr, with a few requiring 5 hr. Students sign up for studies from a list of 10 to 20 available studies per quarter. As an alternative to research participation, students may fulfill their research hours by typing as many as six 1- to 2-page reviews of research articles, with each review equaling 1 research hr. Failure to complete all 6 research hr, in any combination of papers and participation, during the quarter results in an “incomplete” grade that the student must finish during the next quarter or else an “F” grade automatically replaces the incomplete. The instructor changes the incomplete grade to the student’s earned letter grade after the student finishes the requirement.

A staff person hired by the psychology department recorded and maintained research participation hours. The staff person updated official records of earned hours from daily logs submitted by researchers. The staff person also read and logged the article reviews on a daily basis. Students could check a posted list to review their accumulated credits. At the midterm and end of term, the staff person distributed the list to instructors.

Before the onset of the incentive program, completion of the research requirement was relatively low. In the Winter 1998 quarter, 53.8% of the enrolled students failed to complete the requirement and received an incomplete for the course. In the Spring 1999 quarter, 60% of enrolled students failed to complete the research requirement on time. We decided to offer an incentive bonus-point procedure to students in future classes who completed their research hours early in the term. Specifically, on the first day of class, we informed students that if they completed at least 3 of the research hours before the end of the fifth week of the 10-week term, they would receive 3 extra credit points on an exam. Furthermore, if they completed all 6 research hours before the end of
the ninth week of class they would receive another 3 extra credit points on the last exam. Three extra points were a meaningful amount of credit because each exam was composed of approximately 35 points, and 3 points could raise a student’s score by a letter grade.

A total of 590 students (313 women and 193 men; gender was unavailable for 84 students) in one of five introductory psychology classes (from Fall of 1999 to Spring of 2000) comprised this bonus-point condition. We compared the progress and final grades (recorded on a 4-point scale) of these students to those of 480 students (253 women and 157 men; gender was unavailable for 70 students) enrolled in introductory psychology classes in which the instructor did not offer the bonus incentive. Four classes from Winter 1998 and Spring 1999 (before the introduction of the incentive program) and three classes that met during the same period as the incentive classes comprised the nonbonus group. Five different instructors taught the nonbonus classes, including one of the bonus-condition instructors.

Results

In the five bonus classes, the overall completion rate was 60.5%, which was significantly higher than the completion rate for the control classes (54.4%), $\chi^2(1, N = 1,080) = 4.08, p < .05, \phi = .06$. Using logistic regression to control for number of required research credits and individual differences in academic motivation, we compared the completion rate for students in bonus and nonbonus classes. We entered students' grade for the course (as an indicator of academic motivation) and number of credits required into the logistic regression equation at Step 1. At Step 2, we entered the bonus system variable (dummy coded with bonus class students = 1), to determine whether the bonus system independently explained differences in the completion rate. At Step 1, the predictors did help to explain completion of the requirement, $\chi^2(2, N = 406) = 105.46, p < .01, R^2 = .19$; however, the addition of the bonus variable significantly increased the predictive ability of the regression equation, change in $\chi^2(1, N = 406) = 8.33, p < .01, R^2 = .21$ (see Table 1 for the regression weights). Controlling for number of credits and student grade, the predicted probability of completing the requirement was greater for students in the bonus class (probability = .66) than for students in the nonbonus class (probability = .32). To determine whether the bonus system encouraged students to begin work on the requirement earlier, we examined when students began work on the requirement (i.e., earned their first research credit), tracking on a weekly basis. Students in the bonus classes did begin work on the requirement significantly earlier than students in the control group, $t(685) = 7.65, p < .001, d = 0.74$. In the bonus classes, students began working on the requirement during Week 2, on average ($M = 2.78, SD = 1.84$). In the nonbonus classes, students began the requirement, on average, during Week 4 ($M = 4.23, SD = 2.42$). We also examined how many credits students had accumulated at the midterm point (Week 5 in the quarter). For students in the bonus classes, having at least three credits at this point would earn them extra credit points. The results of an ANCOVA, controlling for the number of required research points, indicated that students in the bonus classes had accumulated significantly more research credits ($M = 3.16, SD = 1.97$) than students in the nonbonus classes ($M = 1.70, SD = 1.56$), $F(1, 528) = 47.49, p < .001, d = 0.77$. Consequently, it seems that the first set of bonus points was effective in motivating students to begin work on the requirement.

Finally, we examined final course grade for the two groups. We coded grades on a 4-point scale similar to grade point average. Although students in the bonus classes did have slightly higher grades on average ($M = 2.70, SD = 1.24$) than students in the nonbonus classes ($M = 2.47, SD = 1.34$), this difference was not significant, $t(529) = 1.72, p < .09$. The bonus points did not seem to impact student grades significantly.

Discussion

The use of an incentive program in introductory psychology classes increased research participation rates. Compared to students in classes without the incentive program, students in the incentive classes started work on the research requirement earlier, completed more of the requirement by midterm, and were more likely to finish the requirement during the term in which the class met. For students, the incentive program provided motivation to work on the requirement, thereby decreasing their chances of earning an incomplete grade for the course. In pedagogical terms, early research participation may improve understanding of research methods and aid in understanding of topics presented in class (Holmes, 1967; Landrum & Chastain, 1995). Exposure to typical research procedures may aid the understanding of research findings presented in class. Students may better understand the difference between empirical and library research with firsthand research experience. They may also better appreciate issues faced by research psychologists, such as demand characteristics and deception. The increased knowledge should help students critique research findings. Although this study did not assess any of these possible educational gains, future researchers may do so. Instructors might ask about students’ research experience before quizzes and then look for grade differences between students who have and have not participated in relevant studies (for another method of investigating research studies’ value, see Moreland, 1999). At schools without formal participant pools, this finding is still useful when students may participate in occasional research opportunities. Bonus points may offset students’ hesitancy about expending time participating.

| Table 1. Logistic Regression Weights Predicting Requirement Completion |
|---------------------------|-----------|-----------|
| Predictor                | $B$       | SE of $B$ |
| Number of required credits | 0.79*     | 0.33      |
| Course grade             | -0.99**   | 0.12      |
| Bonus or no bonus        | -0.98**   | 0.35      |

Note. Constant = 2.98. Number of required credits ranged from −5 (5 credits) to +5 (6 credits). Bonus ranged from 0 (no bonus) to 1 (bonus). Requirement completion ranged from 0 (completed) to 1 (not completed).

*p < .05. **p < .001.
Although use of an incentive program may have some drawbacks (e.g., additional work of recording extra credit points), the benefits of increased participation, increased classroom learning, and a more representative sample may balance these drawbacks. By motivating students to participate in research, their education may be enhanced.

References


Notes

1. Portions of this paper were presented at the 2000 meeting of the Eastern Psychological Association, Baltimore.
2. We thank Sue Conwell for data collection and entry.
3. Send correspondence to Joseph R. Ferrari at Department of Psychology, DePaul University, 2219 North Kenmore Ave., Chicago, IL 60614; e-mail: jferrari@depaul.edu.
Using a Dining Facility As an Introductory Psychology Research Laboratory

Nancy Koschmann
Elmira College

Richard Wesp
East Stroudsburg University of Pennsylvania

Because of the large number of students typically enrolled in introductory psychology classes, the use of formal laboratory activities is sometimes impractical, especially when resources are scarce. We describe the use of a college dining facility as a real-life laboratory where students can apply research skills, and we offer evidence that the experience serves as an effective tool to teach the scientific method. The advantages of using this environment are that it includes a rich source of interesting behavior to study and it requires no financial support.

McGovern, Furumoto, Halpern, Kimble, and McKeachie (1991) noted that the most common definitions of psychology refer to it as a science, and Miller (1992) called on teachers of introductory psychology to develop in their students an understanding of and appreciation for the scientific nature of the field. However, most general psychology textbooks provide limited coverage of research concepts early in the text and rarely apply those concepts to the material throughout the remaining chapters (Hedricks, Marvel, & Barrington, 1990). We believe faculty should give introductory psychology students an opportunity to apply research skills, thus involving them in the broader scientific process (Wesp & Koschmann, 1998).

Approaches to teaching the scientific method to introductory psychology students range from complex laboratory requirements to rather simple demonstrations. Examples of more elaborate procedures include Kohn and Brill's (1981) introductory laboratory run by students and Goolkasian and Lee's (1988) computerized laboratory experience. Examples of less complex activities include Stallings's (1993) demonstration that manipulation of amount of fertilizer influences growth of seedlings, Johnson's (1996) evaluation of handedness, and a variety of demonstrations described in teaching handbooks edited by Benjamin and Lowman (1981) and Ware and Johnson (1996).

Some approaches require resources often not available for a large-scale, introductory-level course; others fail to provide comparisons of the various methods available to psychologists. Several years ago we began to use our college dining facility as a real-life laboratory to teach observation skills. This environment offered a wide range of observable behaviors that occur with little or no planning or expense. We expanded our use of the dining hall as we recognized the value of this facility for teaching the various elements of the scientific process, including hypothesis formation, operational definitions, sampling, unobtrusive measures, manipulation of variables, data collection, statistical analysis, and interpretation of data. Research in this setting lends itself to interesting literature searches, discussion of research ethics, and further questions for exploration. We describe some of the ways we used the dining facility as a laboratory to teach scientific methods. Furthermore, we evaluate student perception of the activities and offer evidence of the effectiveness of these active learning experiences.

Research Activities in the Dining Facility

Simple Observation

We first used the dining facility to teach observational skills such as operationally defining variables, sampling, and descriptive statistics. Because the risk to those observed was minimal, the behaviors not of a sensitive nature, the observations unobtrusive, and the setting public, these learning activities were relatively unencumbered by ethical concerns.

Activity. We told students to observe the behavior of others during mealtime in the cafeteria. The instructor gave a few general examples (e.g., seat selection or food choices) but encouraged students to come up with their own behaviors to investigate. We provided no details other than reminding students to record their observations, maintain confidentiality, and explain what they were doing if asked.

Limited instructions led to a wide variety of observations reported at the next scheduled class. Some narrated in detail how people chose a seat, ate, socialized, or departed. Others took a quantitative approach and counted drinks taken or return trips for food. Some tried to interpret emotional responses to such things as finding a seat or meeting a friend. Still others made comparisons between groups based on such characteristics as gender, social status, or college class. This diversity of creative observational approaches provided excellent fodder for discussion of the breadth of the issues studied and methods used by psychologists. Subsequently we
considered refinements of the methods used and made clear that their activities were part of the process of science.

Correlational Study

Once students develop an appreciation for some of the basic observational tools, we introduce them to techniques used for showing relations among variables. As correlational research often requires a more intrusive approach, introduction of the assignment serves as a vehicle to explore ethical responsibilities of researchers. The following are two examples of activities we use to introduce the concept of correlation. Both are useful in demonstrating the process of gathering multiple responses from individuals, graphing data, interpreting correlational statistics, and understanding the limitations of the conclusions one can draw from normative research.

Activities. In a small-group demonstration of correlation, the instructor directed groups of 4 to 6 students to observe two related behaviors they might see in the cafeteria. We offered examples of related behaviors (e.g., the amount of food taken and how far from the food line students sat or number of napkins taken and how many items were unused) but encouraged students to select observations for which they could form a hypothesis describing the relation. The instructor explained that the observable behavior should be quantifiable and gave examples (e.g., count the number of tables away from the food line as a measure of distance). We gave groups 15 min to discuss their ideas and met with each to offer suggestions. We reminded students to use good recording techniques and to observe ethical standards.

In a large-group activity we asked all members of the class to observe the amount of food taken from a salad bar where cost was determined on a container basis and to compare that with another variable. One class discussed means of measuring amount and agreed that because a standard-size container was used, the height of the salad in millimeters was a reasonable measure of the amount of salad served. Students asked diners to fill out a short questionnaire with categories such as age, class in school, sex, and weekly allowance and then asked the diner's permission to measure the height of the salad. Students informed diners of the purpose of the study, told them that their names would not be used, and told them they did not have to submit the survey if they felt at all uncomfortable about it. Another class used a less obtrusive approach, comparing estimates of amount of salad with estimates of height, weight, and age.

In both activities, students reported their findings at the next class. The instructor aided those testing a hypothesis in analysis of their data. In the small-group approach the instructor helped each group plot results on a scatterplot. In the large-group approach each student conducted one observation and plotted his or her results on a scatterplot on the blackboard. The instructor calculated the correlation coefficient in class or between class meetings.

With both activities, students observed the research process from the inception of a hypothesis to drawing conclusions from data. We used the results as a stimulus to discuss problems with causal interpretation drawn from normative data. For example, students believed that a negative correlation between allowance and amount of food taken meant that diners who felt "poor" tended to take more food. Following a discussion of this correlation, students understood that an alternative interpretation was that students with smaller allowances may come from frugal families that have taught the value of getting the most for one's money.

Experiments

Student recognition of the limitations of normative research conclusions provides a natural segue into the discussion of experimental design. We begin the discussion of experimental approaches with further comparison of normative and experimental procedures and conclusions, problems with overgeneralizing conclusions, more detailed discussion of ethics, and choice of appropriate experimental designs.

Activities. One of the simplest experimental approaches employs the participant as his or her own control (i.e., single-subject AB designs and quasi-experimental pretest–posttest designs). In one version of this approach we told students to consider a behavior that occurs during mealtime and to identify an environmental variable that might change that behavior. After students formed hypotheses about what factors might influence behavior, we instructed them to observe and record the behavior they identified, change the environment, and then observe and record changes in the behavior. The instructor gave examples of manipulations (e.g., to study modeling one might measure the frequency of tablemates looking toward one other table, begin to stare at that table, and record the frequency of other tablemates modeling one's staring behavior) and explained the importance of immediately recording observations.

Students conducted simple studies including handing out napkins to assess how napkin availability might increase usage, commenting on manners to evaluate the impact of verbal comments on social standards, and introducing a new friend to a table of diners to evaluate the influence of strangers on social interactions. Students identified elements of their study (e.g., the experimental and control conditions) and we discussed statistical tools, the difference between experimental and other designs, and the importance of replication.

In a group approach, the class met in the cafeteria and designed a study to compare the quality of different foods. Employing a randomized group design we compared brands of yogurts, soft drinks, or pastries supplied by the cafeteria staff. These experiments stimulated discussion of the use of rating scales, the influence of confounds such as testing order, overgeneralizing the conclusions, and alternative methods of design.

Evaluation

In a recent course we evaluated the effectiveness of and student response to active laboratory experiences in a cafeteria. Students in two undergraduate introductory psychology sections taught during the same term by the second author
participated. On the first class day, students completed an 8-item true-false pretest on issues related to the scientific method. The instructor presented the test as an evaluative tool to assess their understanding of the basic concepts of research; he did not review the answers. Questions required students to identify the methods of research and conclusions one might draw from the research.

The instructor assigned research activities in the cafeteria to the 44 students enrolled in one class, whereas he did not assign research activities to the 41 students in the second class. He assigned the first activity described in the previous observational, correlational, and experimental activity sections on the first, second, and third class days, respectively. The instructor asked students to describe their findings in classes subsequent to the assignment and integrated student responses into a discussion and lecture about the different methods of science. Additionally, the instructor assigned students in both classes readings in Myers’s (1996) *Exploring Psychology* and gave lectures about the methods. The instructor focused class lectures and discussion in the second class on examples of research from the psychological literature. On the sixth class day, students took a 20-item exam that included the 8 pretest items and 12 multiple-choice items on other issues studied in the first unit of the course that included material in chapter 1 of their text. The exam was worth 6% of the final course grade.

At the end of the course 37 students in the research assignment course completed an evaluation that asked them to rate how strongly they agreed with statements about the course. Two statements referred to the assignments in the cafeteria. The first posited that the activities assigned in the cafeteria helped them understand the scientific method, the second that the activities helped show how the scientific method applied to psychology. Students rated each statement on a 9-point scale ranging from −4 (very strongly disagree) to 4 (very strongly agree).

**Results**

We analyzed performance on tests of the 42 students in the research assignment class and 39 in the comparison class who completed both the pretests and posttests. An alpha level of .05 was used for all comparisons. A t-test showed no significant difference in the number of errors on the pretest made by the research assignment class (M = 3.31, SD = 1.39) and the comparison class (M = 3.36, SD = 1.46), t(79) = 0.16, p > .05. Therefore, we assumed that the two classes were comparable.

A significantly larger portion of those in the research assignment class (78.6%) improved their exam performance from the pretest to posttest than did those without the research experience (53.8%), χ²(1, N = 81) = 5.56, p < .05.

We used a t-test to compare course-element ratings to a neutral response of 0 on the rating scale. Students agreed (M = 2.11, SD = 2.21) that the activities assigned in the cafeteria helped them understand the scientific method, t(36) = 8.42, p < .05. Furthermore, they agreed (M = 2.38, SD = 1.55) that the activities helped show how the scientific method applied to psychology, t(36) = 9.32, p < .05.

**Discussion**

The research activities stimulated considerable discussion about scientific methods, were reasonably well rated, and appeared to lead to better performance on a methods exam. The findings are not surprising. Benjamin (1991) encouraged the use of active learning procedures in the teaching of psychology, pointing to some successes in his large lecture classes. Others (e.g., Ferraro, 1990; Krehbiel & Lewis, 1994) described research activities that occur outside the normal teaching environment and concluded, as do we, that real-world issues appear to make the study of research more relevant and interesting.

The examples we listed are prototypes for other activities we use. One might adjust assignments to require more or less individual effort or use them singularly rather than in sequence. Students can design their own studies to follow up on class observations. For example, students conducted additional observations focusing on the relation between food-related activities and issues such as participants’ gender, body type, and role in the college community. We also use activities in the cafeteria to teach research concepts in upper level courses.

Although the observational activities we described tend not to require consideration by an institutional review board, others may. One-time approval may be acceptable for activities conducted the same way each time the class is offered. Instructors should engage students in discussion of the ethics of proposed research.

We believe that the study of scientific methods should occur early in an introductory course, but recognize the obstacle of overcoming student perception of science as a “dry” issue. Because of their familiarity with the cafeteria, students seem comfortable engaging in these activities and discussing their ideas. We believe this process accomplishes what Brems (1994) referred to as demystification of the scientific process, an important early step in teaching students about research. These real-life experiences help bring alive the issues surrounding research methodology while setting interactive and engaging expectations for the remainder of the course. The cafeteria provides a rich source of interesting behavior and requires no financial support.

**References**


Notes

1. Portions of this article were presented at the sixteenth annual National Institute on the Teaching of Psychology, St. Petersburg Beach, FL.
2. We thank Linda Noble and three anonymous reviewers for useful suggestions and our students for actively participating in the research activities.
3. Send correspondence to Nancy Koschmann, Department of Psychology, One Park Place, Elmira College, Elmira, NY 14901; e-mail: nkoschmann@elmira.edu.
these questions indicated gains in knowledge (e.g., “gained huge amount of geographical and cultural knowledge”). Students’ comments suggested gains in cross-cultural awareness (e.g., “before I always thought other countries were similar to ours”; “opened my mind to problems throughout the world”). Responses also demonstrated gains in skills through completion of the project (e.g., “experience talking in front of a group”; “how to listen to other students”).

The activity was successful in enhancing social awareness of other cultures and geographical knowledge. Through participation as a presenter and audience member, students explored cross-cultural experiences and became more familiar with research in psychology. Students also suggested that they developed a number of practical skills relating to conducting their research and summarizing it for others.

We have several suggestions to improve the activity. One suggestion is to cover the countries by continent. To present countries in this manner may allow students to see connections among countries in a particular region of the world. Several students suggested, and we agree, that the presentations would be improved by increased visual aids, such as pictures of individuals from the country, a world or country map, or figures of the statistics under question.

Class size may impact the effectiveness of the presentations. Based on our experiences, it is best to limit presentations to one a day. Students were more involved and attentive when there was a single presentation. In large class sections, pairs of students could present on one country.

Depending on the course, instructors can vary the level of guidance they give students in selecting topics. However, students may enjoy the freedom of selecting their own topics. For example, in the course in which the instructor restricted topics, feedback from students indicated that some felt limited by the suggested topic areas. Faculty members should provide guidance on the Internet search options. Some students had difficulty judging the quality of online sources and making the distinction between Web sites and online journals.

A final consideration for faculty members will be how to integrate the class presentations with the particular class topic. In our opinion, daily integration of topics is not necessary for the activity to be meaningful to the course. In fact, on occasions when presentations did not match the lecture, we found that students discussed led to connections across course topics (e.g., how marriage practices and expectations influence career aspirations) and reinforced the underlying themes of the course.

References


Appendix
Suggested Resources for the Around the World Activity


Note

Send correspondence to Gabie E. Smith, Department of Psychology, Campus Box 2163, Elon University, Elon, NC 27244; e-mail: gsmith@elon.edu.

An Argument for a Laboratory in Introductory Psychology

Howard C. Berthold
Christopher M. Hakala
Lycoming College

Dennis Goff
Randolph-Macon Woman’s College

Understanding psychology as a science is essential for the introductory psychology student. Considering that introductory psychology is typically the only course that students take in the discipline, we believe that the course should help students develop an appreciation of the scientific nature of the field. We argue that a laboratory in the introductory psychology course provides an excellent opportunity to emphasize the importance of the scientific method in psychology.

The St. Mary’s Conference on Enhancing the Quality of Undergraduate Education in Psychology prompted several recommendations for more data on current curricular practices (Brewer et al., 1993; McGovern, 1993). Several researchers responded to this call (Messer, 1997; Messer, Griggs, & Jackson, 1999; Perlman & McCann, 1999; Stache, Perlman, McCann, & McFadden, 1994). These initial reports provided guidance and opportunity for discussion on ways to improve the undergraduate curriculum. One aspect
of the available data that surprised us was Perlman and McCann's finding that only 5% of the introductory courses in their sample included laboratories. We believe that a laboratory component in the introductory psychology course benefits both the psychology curriculum and the objectives of the overall undergraduate educational experience.

Psychology in the Undergraduate Curriculum

As far back as 1985, Scheirer and Rogers reported that the vast majority of undergraduates, regardless of major, took introductory psychology. In 1997, McDonald reported a surge in undergraduate psychology majors during the previous decade. The introductory course, therefore, represents the field's best opportunity to portray the discipline to the majority of the college-educated populace in the United States. What do students need to know about psychology?

Brewer et al. (1993) stated "The fundamental goal of education in psychology, from which all the others follow, is to teach students to think as scientists about behavior" (p. 169). Despite widespread support for this position (Freeman, 1997; McGovern & Reich, 1996; N. E. Miller, 1992), a nationwide survey showed that although most instructors of introductory psychology courses stated that "their most important goal is to engage students in scientific inquiry about psychological processes ... few believe they achieve this goal" (B. Miller & Gentile, 1998, p. 95). In a review of 46 widely used texts, Hendricks, Marvel, and Barrington (1990) concluded that Introductory Psychology textbooks often present limited coverage of research methodology. Objective analyses of brief introductory psychology textbooks by Griggs, Jackson, and Napolitano (1994) and Griggs and Koenig (2001) found the mean percentage of text devoted to the topic of methods was 3% for both time periods.

Benefits of an Introductory Psychology Laboratory

A primary benefit of an introductory psychology laboratory, then, is that it goes beyond giving limited lip service to the importance of the scientific approach in psychology. By reading past research; creating hypotheses; developing testing procedures; and collecting, analyzing, and interpreting data, students experience firsthand how some of the principles in their texts came to exist.

Because of the intensive nature of the laboratory, we typically limit laboratory enrollment to 25 students. During laboratories, students and the instructor interact in a less formal atmosphere. Such interaction may not only improve interest, motivation, and attitudes toward the course and the instructor (Billson, 1986; Fiechtnr & Michaelson, 1984; Johnson, Johnson, & Smith, 1991; Lawry, 1990; Light, 1992; Nemerow, 1996), but may also have a positive impact on retention at the college or university (Astin, 1993; Tinto, 1993). Although there are many ways to incorporate active learning into classes, Faust and Paulson (1998) and Mathie et al. (1993) noted that many instructors find it helpful to use easy, low-risk approaches. Laboratories provide a natural structure for blending active learning with hands-on experience in working through stages of the scientific process.

Laboratories also provide a natural structure for writing assignments. Light (1992) found that students value strong writing skills and that learning is more effective when writing is organized around a substantive discipline. B. Miller and Gentile (1998) found that in the typical introductory course writing, if required at all, is limited to a short paper or essay exam. Weekly laboratory reports using American Psychological Association (APA; 2001) style provide considerable opportunity for enhancing writing skills through extensive practice, prompting, and feedback. In addition, the introductory laboratory can specifically improve students' appreciation of the scientific nature of the discipline, performance in an associated lecture course, and preparation for a research methods course.

Self-Report Data

Goff (1998) found that students gave high subjective ratings to six questions relevant to participating in laboratory exercises on end-of-course evaluations. Students reported the laboratory experience increased their understanding of lecture and textbook materials and their interest in the discipline. Students also reported that the active elements of the laboratory were among the most helpful to them as they made connections among the material in laboratory, lectures, and the text. Goolkasian and Lee (1988) made a similar observation about their students' attitudes toward the effectiveness of their laboratory exercises. Furthermore, the first author (Berthold) found that students rated a laboratory as having provided a great deal of help in understanding the scientific method, helped with learning APA writing style, increased their interest in the course, increased the interaction with their professor, and helped the students improve their writing.

Performance Data

In addition to the positive course evaluations, Goff (1998) found that students taking the laboratory had higher final averages on the lecture portion of the course than students who attended the same lectures but did not take the laboratory. Koschmann and Wesp (2001) also demonstrated a positive effect on students' ability to identify methods and draw accurate conclusions from research. In comparing two classes of students with similar scores on a pretest, significantly more introductory students in the class with a laboratory performed better on the same test at the end of the term than did students in a class that involved participation in lectures and discussions on research literature but did not require actual research. The finding confirms Ditman, Macalister, and Payne's (1999) research in which students enrolled in their introductory laboratory showed improved performance in a subsequent research methods course compared to students without an introductory laboratory experience.

Room for Improvement

Perlman and McCann (1999), in comparing their data to principles for quality undergraduate psychology programs (McGovern & Reich, 1996), argued that there is room for
improvement in providing laboratory experiences for undergraduate majors (see also Perlman & McCann, 2002). Additionally, Sternberg (1999) noted “students can and should learn not only about the contents of research, but also about how to generate and evaluate research” (p. 212) and that small research groups should pursue projects “from start to finish” (p. 212). Merely discussing experiments, even in active learning groups, does not provide the same educational opportunity as actually conducting experiments (Goolkasian & Lee, 1988; Westwood, 1988). Moreover, it is important to emphasize and practice research throughout the course (Wesp & Koschmann, 1998), not just in conjunction with the first chapter or two of the introductory textbook, where the topic normally appears (Hendricks et al., 1990). Weekly laboratories insure continued emphasis on the application of research methods during coverage of the various subareas of psychology throughout the semester.

Practical Issues

Incorporating a laboratory into the introductory psychology course does not require expensive laboratory equipment or supplies (Koschmann & Wesp, 2001). There are a number of options for establishing a teaching laboratory in an existing computer laboratory (Goff, 1998; Goolkasian & Lee, 1988). Without this option there are a number of experiments instructors can implement with minimal equipment (e.g., Kohn & Brill, 1981). For example, students can conduct memory experiments with photocopies of stimulus materials and response sheets and can implement a Stroop paradigm with photocopies of stimulus materials and a stopwatch.

The largest cost for the introductory laboratory arises from the need to limit the enrollment. When class sizes are already appropriate, a 2-hr laboratory could be substituted for 1 hr of class each week. With this kind of change the number of contact hours increases only slightly. As B. Miller and Gentile (1998) found that enrollment exceeded 50 students in 43% of introductory classes nationwide, this kind of adjustment will not be possible for many courses. Large classes must be split into smaller sections for laboratories, and those laboratories need to be staffed. The University of Richmond satisfied this need by establishing an introductory laboratory staffed by advanced students (Newcomb & Bagwell, 1997).

Administrators and colleagues often resist curricular changes that require additional resources unless the changes are buttressed by a strong rationale and evidence of success. Beginning modestly and offering a few laboratory sections as optional one-credit supplements to the regular class might satisfy this skepticism. Another option would be to offer one section of an alternative laboratory-based introductory psychology section using the model of dropping 1 hr of class a week for 2 hr of laboratory. A comparison of the performance of students in these optional sections to the performance of students in nonlaboratory sections should demonstrate the advantages of a laboratory in introductory psychology.

Conclusions

An introductory psychology laboratory provides an exceptional opportunity to demonstrate and emphasize the importance of the scientific method in the field of psychology to students of all majors. By participating in all aspects of the scientific process, students learn how to think critically about complex issues. They learn how to write in a clear, cogent manner according to a prescribed format and enjoy the benefits of participating in small, active learning groups. Psychology majors develop specific skills relevant to subsequent course work and the empirical foundations of the discipline. We argue that laboratory work is so integral to the field of psychology and the benefits to majors and nonmajors alike are so important that a laboratory should become the norm rather than the exception in the introductory psychology course.

References


What’s in a Name? Better Letters
If It’s Mine!

Angela Lipsitz
Lance A. Gifford
Northern Kentucky University

We describe a simple classroom demonstration of the name letter effect, the preference for letters in one’s own name (Nuttin, 1985), and provide support for the demonstration’s effectiveness.

Although popular psychology literature suggests that low self-esteem is a widespread problem, research literature indicates that, quite the contrary, most individuals think rather highly of themselves. In fact, people employ a vast arsenal of self-serving cognitions to maintain self-esteem at elevated levels. People rate their traits and abilities as above average (Alicke, 1985) and see the traits and abilities they possess as the ones most important to have (Dunning, Perie, & Story, 1991). People see their successes as caused by these traits and abilities and attribute their failures to other individuals, bad luck, or events beyond their control (Bradley, 1978). In addition, people exaggerate their roles in past events (M. Ross & Sicoly, 1979), overestimate their similarity to others (L. Ross, Greene, & House, 1977), and are unrealistically optimistic about future events (Weinstein, 1980).

So high is self-esteem that people even think the letters in their names are better letters. Nuttin (1985) called this finding the name letter effect and regarded it as a specific case of the mere ownership effect, the idea of valuing objects that are part of oneself more than objects that are not (Heider, 1958; see also, Beggin, 1992). Nuttin (1985, 1987) found that, if asked to choose their favorite letter from two- or three-letter groups or to choose their six favorite letters from a larger group, research participants chose letters in their first and last names at a higher-than-chance frequency. The effect has been demonstrated in over a dozen languages, including those using non-Roman alphabets (Hoorens, Nuttin, Herman, & Pavakanun, 1990; Nuttin, 1987). Researchers have shown that the effect is not due to name letters being more frequent (Nuttin, 1987), to an attachment to letters first written (Hoorens & Todorova, 1988), or to participants guessing the purpose of the research (Nuttin, 1985).

Like many of the other self-serving tendencies, the name letter effect can be demonstrated in class. Boutright-Horowitz (1995) offered one method, but it is cumbersome to carry out, as it necessitates dividing the class into pairs, having half the class leave the room and having students construct test materials. We offer an alternative method, one that takes less time and is suitable for use even in large classes. In this variation, the entire class participates at once, and everyone uses the same already prepared test materials. Furthermore, a show of hands is usually sufficient to demonstrate the effect.

Instructions

Conducting the demonstration requires about 20 min of class time and copies of a handout containing the letters of
The Pedagogical Value of Experimental Participation Paired With Course Content

Michelle Ceynar Rosell, Danielle M. Beck, Katie E. Luther, Kelly M. Goedert, Wendelyn J. Shore, and Dana D. Anderson
Pacific Lutheran University

This study investigated the educational value of research participation by assessing the accuracy of student perceptions regarding the scientific status and methodology of psychology at 3 times during a semester: during the first week, following introductory and methodology lectures, and at the end of the term. Students’ understanding of contemporary psychology and research procedures improved across the term. Findings indicate that students’ increased understanding of psychological research procedures may be due to their participation in research in addition to course content.

Two goals that appear to be germane to most introductory psychology courses are that students gain an understanding of (a) the breadth of contemporary psychology and (b) the scientific methods psychologists employ. College students taking their first psychology course often have misinformed opinions about psychology based on exposure to the popular media. For example, the prevalence of psychological television talk shows and self-help books, in addition to the iconic status of Sigmund Freud as representative of the field of psychology, might lead students to overestimate the extent to which psychology is a clinical field that relies on armchair observation methods (Stanovich, 1986). Often, teaching students to appreciate contemporary psychology necessitates attempting to correct these popular misconceptions (McKeachie, 1960; Vaughan, 1977). Indeed, several instructors explicitly mention on their syllabi that one of the goals of the course is to debunk popular myths regarding psychology (Project Syllabus, 2003), and correcting common opinions based on media misinformation is an avowed objective of popular introductory texts (e.g., Wade & Tavris, 2003) and supplements (e.g., Stanovich, 1986).

Does the common practice of requiring introductory students to participate in research help to meet these pedagogical goals of introductory psychology classes? Sieber and Saks (1989) found that 74% of the universities they surveyed used a participant pool, 93% of which recruited participants from introductory courses. Universities often claim educational value as the rationale for requiring introductory psychology students to participate in experiments (Jung, 1969; Landrum...
Our study assessed the experimental evidence for the educational value of research participation, particularly whether such participation helps debunk common myths regarding psychology.

Many of the previous investigations about research participation have evaluated students’ perceptions of their experiences (e.g., Britton, 1979; Nimmer & Handelsman, 1992). Students do regard participation as an enjoyable experience, but tend to rate the educational value of participation somewhat less favorably (Britton, 1979). Students’ perceptions of the educational value of experiments are affected by whether their participation was mandatory and by what time during the semester they are questioned (Nimmer & Handelsman, 1992). As mentioned by Britton (1979), it is not possible to determine whether students’ opinions about the educational value of research participation are less favorable because the experience actually is of little educational value or because students are unable to wholly appreciate its educational value.

Recently, researchers have assessed student satisfaction along with knowledge gained through experimental participation (as opposed to whether students believe they have learned something). Bowman and Waite (2003) found that students who participated in a research option (i.e., participating as a volunteer in research projects, participating in a “mass testing” session, or writing a brief summary of published research articles) had more positive perceptions of psychology and psychological research. More important, students who participated in research had a better understanding of research procedures than those who did not participate (Bowman & Waite, 2003), suggesting that research participation may increase students’ appreciation of the scientific methods employed by psychologists. As Bowman and Waite suggested, however, the design of their study left doubt as to whether students who participated in research had more knowledge of research procedures prior to their participation in the experiments.

We used a repeated measures design to further investigate Bowman and Waite’s (2003) finding that research participation increased knowledge of research procedures. Furthermore, this study explored whether participating in psychological research helps meet the goals of introductory psychology classes that involve correcting students’ ill-informed opinions about the field of psychology and increasing the accuracy of students’ beliefs regarding the scientific nature of contemporary psychology.

We designed a survey to assess students’ perceptions of the contemporary and scientific nature of psychology. Introductory psychology students completed the survey three times throughout the semester: during the first week (prior to both explicit class instruction on the breadth and scientific rigor of contemporary psychology and research participation), during the fifth week (after explicit class instruction on the scientific nature and research methods of contemporary psychology, but before students participated in actual research), and during the final week of class (after students completed their research requirement).

We hypothesized that students would develop more accurate beliefs about psychology by having direct experiences with psychology as a science (i.e., completing experiments, reading journal articles, or attending colloquia). Because the majority of these experiences were not available to students until after the second administration of our survey, differences in students’ perceptions between the last two testing sessions should reflect knowledge acquired through experiences gained from the research-related activities being completed along with in-class experiences, whereas differences between the first two testing sessions should reflect only in-class experiences.

**Method**

**Participants**

Two-hundred twelve students (66 men, 146 women) enrolled in one of three sections of an introductory psychology course at a small private university completed the questionnaire at each of three testing sessions and consented to have their data included in the study. The majority of participants were first-year students (77.2%); 16.6% of participants were sophomores, 3.8% were juniors, and 2.4% were seniors. The age range of participants was 17 to 42 years (M = 18.65, SD = 2.14). Students reported a mean high school grade point average of 3.6 (SD = .33). Forty-eight percent reported taking the class for their intended major. During the first week of class, 23.1% of the students reported participation in a previous psychology study, and 6.6% reported they had read a professional psychological journal.

**Materials**

Students indicated their perceptions regarding the field of psychology on a 43-item questionnaire. We developed items based on those student misconceptions we had commonly encountered in our teaching. Questions focused on ethical practices in psychological research (e.g., A psychology experimenter must receive written consent from the participants in his or her experiment), appreciation of the breadth of contemporary psychology research (e.g., A lot of psychological research can be applied in education or business), and common myths regarding the exclusively clinical nature of psychology (e.g., Most psychologists do research on mental illness). Students rated their level of agreement with each statement using a 6-point Likert-type scale ranging from 1 (strong agreement) to 6 (strong disagreement). This type of scale offered a more sensitive measure of the students’ perceptions about psychology than would a dichotomous scale. Thus, we took the degree of agreement with a factually accurate statement about the nature of psychology (and conversely disagreement with a factually inaccurate statement) to indicate better informed opinions than did less extreme ratings. A 6-point scale avoids neutral ratings.

In addition to demographic items, the students indicated how many research participation credits they had fulfilled by participating in research, summarizing colloquia, or summarizing journal articles. Four open-ended questions added to the second and third administrations of the test assessed students’ experiences while participating in experiments throughout the semester.
Procedure

Introductory psychology students fulfilled an eight-credit research familiarization requirement intended to supplement the course content with more direct experience with psychological research. Students earned credits by participating in psychology experiments (one credit for every half hour) or by writing summaries of designated journal articles or psychology colloquia (one credit for each summary). If a student failed to complete the eight credits, her or his final grade for the course was lowered one letter grade. Students who completed their remaining number of credits during the following academic term had their grade restored.

Students completed the questionnaire during class at three times during the semester: Week 1 (pretest), Week 5 (Posttest 1), and Week 15 (Posttest 2). Students completed the pretest in each class before material pertaining to research methods was presented and immediately following the explanation of the research familiarization requirement. Two-hundred ninety students were present on the day of the pretest in each class before material pertaining to research methods was presented and immediately following the explanation of the research familiarization requirement.

Results

The students earned the vast majority of their eight required research familiarization credits through participation in experiments (M = 7.72, SD = 1.39 credits). Nearly every student (99.5%) participated in at least one experiment.

The results focus on participants’ responses to the closed-ended survey questions. Researchers coded items such that stronger agreement with a statement indicates a more informed opinion about the issue raised.

Data Reduction

A factor analysis of students’ responses to the 31 objective items on the final posttest questionnaire with a varimax rotation and using minimum eigenvalues of 1.0 revealed four reliable factors, accounting for 40.7% of the variance. Two factors pertained to an understanding of contemporary psychology and two factors pertained to an understanding of research procedures. Researchers selected items with a good factor loading (greater than .55) for inclusion for each factor (Comrey & Lee, 1992).

Perceptions about contemporary psychology. The factor analysis revealed two factors relating to students’ understanding of the focus of contemporary psychology. The first factor (factor loadings range from .59 to .76; Cronbach’s α = .81) included six items regarding psychology’s research focus (e.g., “Psychology can study how groups function,” and “Currently, psychologists must adhere to strict ethical guidelines”).

The second factor (factor loadings range from .65 to .82; Cronbach’s α = .73) included three items dealing with the myth that psychology is strictly clinical in its focus (“Most psychologists do research on mental illness,” “Most psychologists counsel their clients,” and “Most psychologists are therapists”).

Awareness of ethical research procedures. The factor analysis revealed two additional factors pertaining to students’ awareness of ethical research procedures. One factor contained three items regarding knowledge of participants’ rights (factor loadings range from .58 to .80; Cronbach’s α = .70). Items that loaded on this factor were the statements, “Participants have the right to find out the results of the psychology experiment,” “Participants have the right to leave the experiment at any time,” and “At the end of an experiment a participant has the right to find out about the nature of the experiment.”

Three items loaded on a second factor (factor loadings range from .56 to .78; Cronbach’s α = .64). Although the reliability for this factor was somewhat lower than the other factors, the content of the items was clearly related to a single construct, and measures of internal consistency are not necessarily indicative of the unidimensionality of a scale (see Clark & Watson, 1995). The items that loaded on this fourth factor targeted students’ awareness about informed consent (i.e., “When a person participates in a psychology experiment they must be informed about the general nature of the study,” “A psychology experimenter must receive written consent from the participants in his/her experiment,” and “Currently psychologists can run experiments on people without their consent”).

Changes in Students’ Perceptions Across Time

We calculated participants’ scores for the four factors by summing their responses to the items that loaded on that factor. We conducted separate repeated measures ANOVAs on each of the factors with time of test (pretest, Posttest 1, and Posttest 2) as a within-subjects variable. Additionally, we conducted two planned comparisons (pretest vs. Posttest 1 and Posttest 1 vs. Posttest 2) for each factor using paired t-tests. Using the Bonferroni correction to adjust for the family-wise error, we accepted a p value less than or equal to .008 as significant. We predicted that participants’ scores on each factor would decrease from pretest to Posttest 1 (reflecting knowledge gained from explicit coverage in class lecture) and
from Posttest 1 to Posttest 2 (reflecting knowledge gained from the research familiarization requirement in addition to class lecture).

**Perceptions of contemporary psychology.** The within-subjects ANOVAs revealed significant differences in both factors relating to students’ perceptions of contemporary psychology as a function of time. More specifically, students’ understanding that psychology is research focused improved during the beginning of the semester, F(2, 400) = 23.64, p < .0001. Planned comparisons revealed significant increases in the accuracy of perceptions from the pretest (M = 12.29, SD = 3.49) to Posttest 1 (M = 11.26, SD = 3.62), t(204) = 4.58, p < .0001, but not from Posttest 1 (M = 11.26, SD = 3.62) to Posttest 2 (M = 10.72, SD = 3.75), t(202) = 2.14, p = .034.

Students’ awareness that psychology is not strictly clinically oriented also increased throughout the semester, F(2, 400) = 18.32, p < .0001. Planned comparisons revealed significant increases from the pretest (M = 9.45, SD = 2.70) to Posttest 1 (M = 8.92, SD = 3.10), t(206) = 2.81, p < .005, and from Posttest 1 (M = 8.92, SD = 3.10) to Posttest 2 (M = 8.37, SD = 2.84), t(202) = 2.87, p < .005.

**Awareness of Ethical Research Procedures.** The within-subjects ANOVAs revealed an overall significant difference in students’ knowledge about the rights of research participants, F(2, 412) = 36.97, p < .0001, and informed consent, F(2, 400) = 21.33, p < .0001, as a function of the time of test. Planned comparisons revealed students’ knowledge of participants’ rights was not significantly different between the pretest (M = 6.19, SD = 2.32) and Posttest 1 (M = 5.98, SD = 2.59), t(209) = 1.30, p = .20, but their awareness of participants’ rights increased significantly from Posttest 1 (M = 5.98, SD = 2.59) to Posttest 2 (M = 4.86, SD = 2.18), t(207) = 7.09, p < .0001. Likewise, planned comparisons revealed students’ understanding of informed consent procedures increased significantly from Posttest 1 (M = 7.59, SD = 3.17) to Posttest 2 (M = 6.30, SD = 2.87), t(205) = 6.36, p < .0001, but not from the pretest (M = 7.64, SD = 2.67) to Posttest 1 (M = 7.59, SD = 3.17), t(204) = .10, p = .92.

**Discussion**

We designed a study to identify the pedagogical value of participating in psychological experiments. We predicted that experience with research in addition to explicit course instruction would, throughout the semester, increase student knowledge as indicated by the accuracy of students’ opinions about psychology.

**Perception of Contemporary Psychology**

As hypothesized, students’ perceptions about contemporary psychology were more accurate as a result of class content (as demonstrated by the differences between the pretest and Posttest 1). During the first few weeks of class, students heard about the different fields, research interests, and applications of psychology. Class instruction during the first weeks of the semester covered content beginning with the roots of psychology and progressing to psychology’s scientific status and methodology today. During this time instructors attempted to debunk common “pop psych” myths about psychology.

Students’ research participation (displayed by the difference between Posttest 1 and Posttest 2 ratings) led to further increase in knowledge. Students’ perceptions of psychology as a science became significantly more accurate after participation in research, writing summaries of published research articles, and attending colloquia. Apparently, hands-on experience with research complemented students’ understanding of the function, purpose, methods, and fields of psychology.

**Awareness of Ethical Research Procedures**

Results indicated that students’ understanding of participants’ rights and informed consent did not improve from the first week to the fifth week of class despite the fact that instructors explicitly covered the methodology and ethics of psychological research during this time. Course content focused on research with both nonhuman and human participants, classic studies that violated human participant rights, and an overview of the ethical principles to which psychologists strictly adhere. Consistent with earlier studies (McKeachie, 1960; Vaughan, 1977), class lecture alone was apparently not sufficient to significantly change students’ misconceptions pertaining to ethical procedures.

After students had the opportunity to participate in psychological studies, however, their knowledge about participants’ rights and informed consent procedures did increase significantly (between the fifth and final weeks of class). These results suggest that increased exposure to research, in addition to class lecture and readings, contributed to students’ increased understanding of ethical procedures used by psychologists. Anecdotally, the fact that 5 of our students did not consent to have their data used at the end of the semester (an unusual occurrence in research) might show that they not only learned their rights in theory, but understood them enough to be comfortable exercising those rights.

**Conclusions**

Although many universities tout the pedagogical benefit of requiring research participation on the part of introductory psychology students, researchers are just beginning to explore the real educational value of participation (Bowman & Waite, 2003; Landrum & Chastain, 1995). Consistent with previous research indicating that students demonstrate increased understanding of psychology after research participation (e.g., Britton, 1979; Nimmer & Handelsman, 1992), our results provide evidence that participation can aid in meeting at least some of the pedagogical goals of an introductory psychology course. Experience with research colloquia, classic psychological literature, and, predominantly, participating in experiments, appears to lead to more accurate perceptions concerning the breadth and scientific rigor of psychology. Students obtained an increased awareness of current psychological research procedures, specifically informed consent
and participant rights. This gain in knowledge occurred sometime after class lectures about these same topics. Thus, this study provides further empirical justification for the use of introductory psychology participant pools as pedagogical tools.

In this study we were interested in assessing the benefit of a research familiarization requirement in addition to traditionally, instructors of the classes included in this study believed it would be unethical to relinquish the research requirement for a portion of the introductory students, given that the instructors thought the research requirement had educational benefits. Therefore, we did not strictly assess the isolated effect of experimental participation on changes in students’ perceptions of the field of psychology. The lack of true control group limits the causal interpretations we can draw from this work. It seems reasonable to conclude, however, that differences in students’ knowledge about psychology from Posttest 1 to Posttest 2 were due in large part to participation in research. Future research may address these issues by using a control group from an introductory class in a different discipline or conducting the experiment outside the regular classroom setting.

In conclusion, like other active learning methods (e.g., Benjamin, 1991; Mathie et al., 1993), research participation seems to have been an effective means of increasing knowledge about psychology. Direct, hands-on experience with research appears not only to be an enjoyable experience (Britton, 1979), but also an educationally valuable experience for introductory psychology students. Participation in a research familiarization requirement can be an active learning tool that enhances students’ understanding of psychology as a science.

References


Notes

1. Danielle M. Beck is now a graduate student at the University of Washington, Seattle. Katie E. Luther is now a graduate student in the Department of Sociology, University of California, Riverside. Dana Anderson, Kelly Goedert, and Wendelyn Shore contributed equally to this research and preparation of the article.

2. This research was presented at the 2002 Western Psychological Association Conference, Irvine, CA.

3. We thank Elisabeth Pynn and Sara Sanders for their assistance with data entry and the anonymous reviewers for their helpful comments on an earlier draft.

4. Send correspondence to Michelle Ceynar Rosell, Department of Psychology, Pacific Lutheran University, Tacoma, WA 98447; e-mail: rosellmc@plu.edu.
Introductory Psychology Students’ Perceptions of Alternatives to Research Participation

David Trafimow, Laura Madson, and Iola Gwizdowski
New Mexico State University

According to the American Psychological Association ethical guidelines and federal regulations, psychology departments that require introductory psychology students to participate in research studies must also provide students with equitable alternatives to research participation, such as writing a paper. However, most students opt to participate in studies anyway. A possible explanation is that students may perceive writing a paper as a coercive alternative. In contradiction to this explanation, the vast majority of our participants indicated, when asked to assume that the paper and the experiment would be equally effortful, that they would choose the experiment anyway.

Most psychology departments at doctoral-granting institutions require or strongly encourage introductory psychology students to participate in research studies as part of the course requirement (Rosell et al., 2005). A common rationale for requiring or encouraging students to participate in research studies is that students will learn more about psychology and research methods by active involvement in the process than by reading about it in a textbook.

Despite any potential educational benefit to students, the American Psychological Association (APA) mandates that “when research participation is a course requirement or an opportunity for extra credit, the prospective participant is given the choice of equitable alternative activities” (APA, 2002, p. 12). Most departments use the alternative of writing a paper rather than participating in research. Some departments offer extra credit or bonus exam points in their introductory psychology course in exchange for participation (Ferrari & McGowan, 2002; Padilla-Walker, Zamboanga, Thompson, & Schmersal, 2005; Sieber & Saks, 1989).

Despite the available alternatives, the majority of students choose to participate in research (Rosell et al., 2005). One interpretation of this pattern is that the vast majority of students honestly prefer being actively involved in the research process. Students might find studies more engaging, more interesting, or more educational than writing papers or some other alternative. Consistent with this interpretation, previous research suggests that students perceive “a modest positive impact” of research participation (Britton, 1987, p. 269; see also Bowman & Waite, 2003; Britton, Richardson, Smith, & Hamilton, 1983; Landrum & Chastain, 1995; Leak, 1981; Nimmer & Handelsman, 1992). (An exception may be when students are not debriefed properly; Coulter, 1986.) In addition, students’ general perceptions of their research experience tend to be more positive than their specific perceptions of educational value received (Britton, 1979; Landrum & Chastain, 1995; Nimmer & Handelsman, 1992). However, when studies measure changes in students’ knowledge as a result of research participation rather than merely measuring their perceptions of the experience, research participation does appear to have some educational value in that students knew more about research procedures following participation in research than they did prior to their participation (Rosell et al., 2005).

On the other hand, students may choose to participate in research because they perceive the proffered alternatives to experimental participation to be sufficiently noxious that the alternatives are, in fact, coercive. Sieber and Saks (1989) suggested that there are typically too few alternatives to experimental participation and that students perceive these alternatives to be effortful or time consuming. They concluded that “limited and unattractive alternatives constitute an element of coercion and belie the educational benefits that are supposed to flow from the research participation requirement” (p. 1057). If the majority of students participate in research studies because the alternative is more time consuming or effortful, it would be a serious violation of ethical guidelines articulated by the APA, various institutional review boards, and many federal granting agencies (APA, 2002; Korn, 1988; Porter, 1991; Reatig, 1982). It may also mean that justifying experimental participation in terms of its educational benefit is a fraudulent use of course requirements.

Despite the importance of this question, little research has explored why students choose to participate in research studies rather than completing other alternatives. Researchers have examined related questions, such as whether students are more satisfied after participating in research or writing a research paper (Bowman & Waite, 2003; Waite & Bowman, 1999) but these studies did not explore why students chose research participation in the first place. Leak (1981) asked students whether offering extra credit in a course in exchange for research participation was coercive. Interestingly, students reported that extra credit was coercive, although they did not find the coercion to be objectionable. Possibly, students find the extra credit to be sufficiently positive as to outweigh the negativity of being coerced. Alternatively, students may not have understood the term coercion. We designed this study to ascertain students’ perceptions of
experimental participation and its alternatives and their reasons for selecting participation over those alternatives.

Method

Participants

We surveyed 125 participants who were enrolled in an introductory psychology course at New Mexico State University. Students participated in approximately 4 hr of research studies during the term. They had the option of summarizing empirical journal articles rather than participating in studies, although less than 1% of them wrote the papers. We collected the data anonymously, and all participants completed an informed consent statement specifying the voluntary nature of participation prior to taking the survey.

Materials

Students responded to an eight-item survey during regularly scheduled class time. The idea of the survey was to assess students’ perceptions about participating in research; their perceptions of the alternative of writing a paper; and, most important, their perceptions of these tasks under conditions of equal time and difficulty. Completing the survey required 5 to 10 min.

Results

The variables of interest were students’ choices regarding experimental participation and their perceptions of the available alternative. Consistent with previous research, students had somewhat positive perceptions of research participation. A substantial minority (43.5%) of students became more interested in psychology as a result of their research participation although the majority of students (51.6%) indicated that research participation had no effect on their interest in psychology. Similarly, 81.6% of students indicated that their research participation was at least slightly enjoyable, and most (84.8%) thought research participation was no more than a slight inconvenience. Almost half (49.6%) of students indicated that they would have learned more from spending an equal amount of time in class.

When asked why they chose research participation instead of writing a paper, students’ perceptions were mixed. Many participants chose the experiment for positive reasons: 43.2% were curious about the experiment, 48.8% thought the experiment would be more interesting, and 20.0% thought they would learn more from the experiment than from writing a paper. On the other hand, 58.4% of students indicated that they thought the paper would take more time than participating in research, and 52.0% thought that the paper would take more effort (students could endorse more than one response). These data suggest that many students perceived writing a paper to be a coercive alternative to participating in an experiment. When asked to compare research participation to spending an equal amount of time in class, responses were somewhat mixed. A large minority of students felt neutral about the comparison (43.2%), although most of the remaining 56.8% of students felt positive (87.3%) rather than negative (13.7%) about research participation. When students compared participating in the experiments to writing research papers, their perceptions were more polarized in that only 19.2% of students felt neutral and, of the remainder, 94.4% felt positive about research participation and 5.6% felt negative.

Item 3 was the most important item because it asked participants to indicate whether they would choose research participation or the alternative assuming that the two options would require equivalent time and effort. Consistent with the notion that participation in research confers benefits, 87.9% indicated that they would choose the experiment even under this assumption. The primary motivations were to learn more (15.3%), satisfy curiosity (17.7%), indulge an interest (41.9%), or for an implicit reason (12.9%).

Discussion

Students perceived research participation to be more positive than writing the alternative paper. Slightly more than half of the participants thought that the paper would be more effortful and more time consuming than the experiment, which supports the notion that this particular alternative to research participation might involve some level of coercion. However, when asked what they would choose assuming that the paper and the experiment would be equally effortful and time consuming, the vast majority of students chose the experiment anyway. Furthermore, the vast majority of participants either felt neutral or positive about the experiments compared to spending time in class or writing research papers.

Of course, this study has important limitations. First, although writing a paper about psychological research is a commonly offered alternative to research participation (Sieber & Saks, 1989), it is only one of many alternatives offered nationwide. Additional research is needed to explore students’ perceptions of other available alternatives and why students consistently tend to select research participation over the alternatives (Rosell et al., 2005). Also, we collected the data using introductory psychology students. Students in upper division courses (e.g., research methods) who are also required to earn research credits may have different perceptions of research participation and the alternatives. In particular, more advanced students may derive greater educational value from their research participation than introductory students because more advanced students have a richer knowledge base with which to comprehend their participation experiences. On the other hand, this situation might be sufficiently rare so as to not matter very much.
All in all, requiring students to participate in research appears to confer modest benefits (or at least does not appear to cause harm). Most students derive some educational benefit from participating in research, report either positive or neutral reactions to their research experience, and would still choose to participate in research even when presented with an alternative that would require equal time and effort. However, if psychology departments are going to continue justifying required research participation in terms of educational benefit, perhaps researchers should strive to improve students’ research experiences. Ideally, students would report more than a modest educational benefit, and most students would report feeling positive rather than merely neutral about their research participation.

References


Note

Send correspondence to David Trafimow, New Mexico State University, Department of Psychology, MSC 3452, PO Box 30001, Las Cruces, NM 88003–8001; e-mail: trafimow@crl.nmsu.edu.

Encouraging Distributed Study: A Classroom Experiment on the Spacing Effect

William R. Balch
The Pennsylvania State University, Altoona

Two introductory psychology classes (N = 145) participated in a counterbalanced classroom experiment that demonstrated the spacing effect and, by analogy, the benefits of distributed study. After hearing words presented twice in either a massed or distributed manner, participants recalled the words and scored their recall protocols, reliably remembering more distributed than massed words. Posttest scores on a multiple-choice quiz covering points illustrated by the experiment averaged about twice the comparable pretest scores, indicating the effectiveness of the exercise in conveying content. Students’ subjective ratings suggested that the experiment helped convince them of the benefits of distributed study.

The spacing effect occurs when distributed study results in better memory than massed study does (e.g., Dempster, 1988). This effect holds true for both the spacing of items in a single list (e.g., Underwood, 1970) and the spacing of practice trials (e.g., Baddeley & Longman, 1978). Moreover, the effect applies to both short items like words (e.g., Verkoeijen, Rikers, & Schmidt, 2004) as well as more complex material like textbook passages (e.g., Kraft & Jenkins, 1981) and classroom lessons (Seabrook, Brown, & Solity, 2005).

By frequent testing, an instructor can effectively help students distribute their study and improve their grades (Fulkerson & Martin, 1981). A complementary method might be to use exercises illustrating the benefits of distributed study to motivate students to adopt this strategy by
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
Introducing Students to Psychological Research: General Psychology as a Laboratory Course

Thomas J. Thieman, E. Gil Clary, Andrea M. Olson, Rachel C. Dauner, and Erin E. Ring
College of St. Catherine

For 6 years, we have offered an integrated weekly laboratory focusing on research methods as part of our general psychology course. Through self-report measures and controlled comparisons, we found that laboratory projects significantly increase students’ knowledge and comfort level with scientific approaches and concepts, sustain interest in psychology, and increase critical thinking about psychological research. Implementing a laboratory component in the introductory course increases students’ scientific literacy, reinforces psychology’s claim to scientific status, encourages active learning, promotes quantitative reasoning, and benefits multiple constituencies.

Undergraduate psychology teachers report that among the most important goals for the introductory psychology course are engaging students in scientific inquiry and convincing them that psychology is a science (Miller & Gentile, 1998). The American Psychological Association (APA) Task Force’s rubric for assessing psychology curricula (Halonen et al., 2003) is based on the premise that “the ability to think about and identify psychological problems from an empirical perspective” (p. 197) is the discipline’s central paradigm. How are these goals best achieved? Hailstorks (2007) recently noted that psychology faculty focus considerable attention on the content of the introductory course, but devote scant effort to discussing how to teach its research foundation. She argued that “we simply need to put the science of psychology in the forefront of our teaching” (Hailstorks, 2007, p. 23).

Psychology teachers overwhelmingly acknowledge the need to promote scientific literacy in the general population (Elmes, 2002), and the movement to increase research-related activities in the undergraduate psychology curriculum is well documented (Kierniesky, 2005). However, few departments have modified their introductory course to initiate this process. Perlman and McCann’s (1999) review of 500 college catalogs revealed that only 5% of introductory psychology courses included a laboratory component. Because general psychology is most students’ sole collegiate exposure to the discipline, their involvement in psychological research typically is limited to classroom demonstrations and out-of-course extra-credit experiments. The latter may increase students’ understanding of scientific methodology (Bowman & Waite, 2003; Rosell et al., 2005), but it falls short of fully engaging students in the research enterprise.

For students who do not already possess a scientific orientation, including a research component in the introductory course provides a potential method for improving their understanding of scientific psychology. Berthold, Hakala, and Goff (2003) summarized arguments for integrating laboratory work into introductory psychology and making this configuration the norm, as is the case with first college courses in biology, chemistry, and physics. In the past, architects of introductory psychology courses with integrated laboratories (Brothen, 1984; Goolkasian & Lee, 1988; Kazmerski & Blasko, 1999; Kohn & Brill, 1981; Koschmann & Wesp, 2001; Murray, Pasternack, & Rowe, 1972; Rowland, Kaarianen, & Houtsmuller, 2000) offered evidence of enhanced student achievement in their research-based curricula. By involving students in empirical research at the introductory level, instructors encourage active learning (Benjamin, 1991), promote quantitative reasoning (Lutsky, 2006), and prompt discussion of related topics such as research ethics (Balch, 2006). They also provide opportunities for

In this article, we report evidence from the first 6 years of a curriculum revision project that demonstrates how the inclusion of a laboratory component in general psychology produces long-term positive outcomes in students’ self-perceptions of scientific literacy and learning, is feasible, and benefits multiple constituencies. In Fall 2001, funded by a National Science Foundation Course, Curriculum, and Laboratory Improvement Grant, the College of St. Catherine psychology faculty added a comprehensive, semester-long laboratory to our general psychology course. This laboratory replaced a 3-week laboratory addendum to general psychology that had been operational for more than 25 years. We focused the new curriculum on research by engaging students in various methods of scientific inquiry.

The Laboratory Experience

Project Overview

The primary objective of our general psychology laboratory is to present a unified approach to psychological science by examining five research strategies: literature review, structured observation, experimentation, correlational research, and analysis of archival data. We presented these research tools as means to answer important questions. Students met weekly in laboratory sections taught by upper division psychology majors, hereafter referred to as laboratory instructors (LIs). We presented each method for two or three sessions, culminating in reports written in APA (2001) style. Each student submitted his or her own report. The first was a literature review; the second, third, and fourth were partial lab reports (e.g., only Method and Results); and the final report was a complete manuscript. During the first year of implementation, we required students to write six full-length APA-style reports, which overwhelmed many students. We carefully reshaped the writing requirements, which now serve to build successively on one another. We were also intentional about integrating each project with the lecture-based material through Microsoft PowerPoint presentations and classroom discussions of classic empirical studies (Hock, 2002). We wanted to avoid the perception that laboratory and lecture were separate from one another.

Course instructors evaluated and graded the first and last reports; LIs graded the intervening papers. We encouraged students to revise their reports for an improved grade if the course instructor or LI noted deficiencies on the first draft. Performance on the laboratory reports, lab participation, and a research methods portion of the final exam accounted for 40% of a student’s course grade.

Logistics

Approximately 250 St. Catherine day students register for the general psychology course each year. Lectures taught by faculty (class size M = 28) meet two or three times a week and labs led by LIs (section size M = 15) convene once a week for 75 min (originally 60 min). Lab rosters were not linked with lecture rosters (i.e., in any given lab, there can be students from any of the various lecture sections). Our laboratory space includes eight individual cubicles with networked desktop computers surrounding a staging area with conference-table seating for 16 students.

We wrote a laboratory manual with lecture and lab schedules, APA formatting guidelines, sample reports for a literature review and empirical study, instructions for accessing PsycINFO and creating tables and figures in Microsoft Word, and exercise instructions and report checklists for each of five projects. Throughout the term, we posted updates and ancillary materials to a companion Blackboard site.

Laboratory instructorships are paid positions. Psychology majors participate in a competitive hiring process for these positions. In terms of responsibility and structure, being an LI is similar to what many graduate students experience. Course instructors met with LIs in weekly group meetings to review the week’s project, discuss how to introduce topics and answer students’ questions, and provide feedback on grading student papers.

Five Research Strategies

Literature review. Students began by investigating a research question of their choice in PsycINFO within a constrained domain (e.g., nature–nurture interactions). Citing at least eight abstracts related to a well-defined topic, students wrote a literature review in which they stated their research question, summarized findings, and noted implications for future research. Students learned that abstracts vary widely in detail and readability. We invited them to read the full-text articles, although it was not a requirement. Because students appended the abstracts to their reports, we...
used this opportunity to instruct them on paraphrasing and how to avoid plagiarism.

**Structured observation.** In the structured observation project, we instructed students how to objectively score a specific set of exploratory behaviors exhibited by 3- to 5-year-olds who we had videotaped interacting with the Banta Curiosity Box (Banta, 1970) in the college's laboratory Montessori School. Students used standardized coding sheets for recording observations (exploration time, number of items manipulated, number of questions asked). Paired students computed interrater reliability coefficients for the number of items each child manipulated in this free-play situation. We discussed how to operationalize complex variables such as curiosity and how to differentiate behavioral description from inference.

**Experimentation.** For the first 2 years, we relied on two commercially available computer software packages to introduce experimentation: a single-subject design comparing fixed and variable ratio partial reinforcement schedules using Sniffy the Virtual Rat (Alloway, Wilson, Graham, & Krames, 2000) and an independent groups comparison of free and cued recall of categorized word lists using Psychology on a Disc (Shimoff, Catania, & Matthews, 2000). In Fall 2003, we replaced these two exercises with a within-subjects verbal memory experiment to give students control of designing their own experiments. Students in paired laboratory sections designed experiments to test one another by selecting a memory phenomenon (e.g., isolation, false memory, serial position, word frequency, chunking, imageability effects), test materials, presentation order, and rate. Computers managed presentation and record keeping. In successive weeks, students designed a study, served as participants in their paired group's study, and analyzed and interpreted their study's data. In reaching consensus about their method, students debated the pros and cons of various design features and learned the importance of programmatic research.

**Correlational research.** For the fourth project, students measured each others' heart and respiration rates using iWorx physiological recording equipment (CB Sciences, 2000) under baseline and mild stressor conditions (to generate a stress reactivity measure) and correlated them with students' scores on a college-normed stress inventory (Renner & Mackin, 1998). They learned the difficulties of working in a domain with high individual differences and the rudiments of calculating and interpreting correlations. In 2007, we dropped this exercise and now address correlation in the experimentation and archival methods labs.

**Archival methods.** In the final project, students explored correlations among beliefs, self-perceptions, and self-reported behaviors in an archival data set of responses from first-year St. Catherine students to the Cooperative Institutional Research Program Freshman Survey (cf. Higher Education Research Institute, 2008). Students generated and tested hypotheses using this local sample and related their findings to published empirical studies abstracted in PsycINFO. In different terms, we prompted students to test the relation between self-reported attitudes and behaviors (e.g., do respondents’ attitudes about legalization of marijuana match their pattern of self-reported alcohol and drug use?) and test for cohort differences (e.g., over years, have respondents changed in the importance attributed to goals such as being a community leader, being well off financially, or developing a personal philosophy of life?). The data set included responses to 135 variables from more than 8,000 students across 25 years.

**Laboratory Evaluation Methods**

For 3 years (6 terms, 24 lecture sections, 54 laboratory sections, more than 700 students), we assessed student learning and attitudes toward the new curriculum using three procedures. First, students completed an attitude and knowledge survey, once during the first week of the term and again during the final week, providing self-ratings of their comfort level, knowledge or experience, and interest in 12 content and skill areas targeted by the laboratories. To ensure confidentiality, we matched pre–post responses by numeric code. Second, we administered end-of-term laboratory evaluations. The third procedure focused on actual learning. Students completed a critical thinking quiz adapted from Lawson (1999) during the first week of the term and an alternate form on the final exam. Approximately equal numbers of students completed each form of the quiz at the beginning of the term. Students read six scenarios per quiz that described research situations and attempted to identify a methodological flaw (e.g., lack of a control group, confounded variables, failure to consider base rates, confusion of causation and correlation).

We also tested for long-term impact on student learning. First, we tested a large number of students within the first week of the lifespan developmental
psychology course with a new set of four flawed research scenarios. At St. Catherine, this course is usually the next psychology course students complete after general psychology and is part of the required supporting coursework for students majoring in education, nursing, and social work. We constructed two groups of students based on whether they had previously completed general psychology with a laboratory (n = 39; 32 at College of St. Catherine and 7 elsewhere) or without a lab (n = 37) and balanced them for mean lag time between the courses (M = 2.19 and 2.08 terms, respectively) and course grades in general psychology (M = 3.39 and 3.38 grade points, respectively, on a scale where A = 4; both ps > .05).

Finally, the first author compared performance on the first writing assignment (a literature review) in his two advanced laboratory courses (learning principles; memory and cognition) for students who completed the St. Catherine general psychology course (n = 33) to those who completed this course without lab elsewhere (n = 39). He considered the scores of only those for whom the course was their first advanced psychology research course.

Based on student feedback, the course instructors initiated some curricular changes but the general pattern of student responses remained consistent across terms. Winter term students expressed slightly higher comfort and knowledge of research skills at the beginning of the term than fall students, and they rated the performance of the LIs slightly higher. For the first three analyses, we describe assessment data from 2003–2004 when students completed five projects.

**Results**

**Student Self-Perceptions**

**Pre–post attitude and knowledge survey.** On a 5-point scale ranging from 1 (very low) to 5 (very high), students’ mean entry levels of comfort and knowledge on 12 laboratory-related skills ranged from 1.26 (statistical analysis knowledge) to 4.11 (word processing knowledge). At the end of the term, all the ratings averaged from 2.80 to 4.45 (see Table 1). All 24 within-subjects comparisons yielded significant t ratios indicating higher exit means (all p < .001), even when we applied a Bonferroni correction. Most notably, comfort and knowledge with statistical analysis and database searching increased approximately 2 scale points.

Interest began moderate to high (means ranging from 2.48–3.78) and generally stayed in the same range. There were no significant changes in 8 of the 12 comparisons, whereas 4 showed small but above-chance declines (computer software and equipment skills, understanding and evaluating psychological research, and drawing conclusions from a psychological study). Only interest declines in evaluating research and using computer equipment were statistically significant with the Bonferroni correction.

**End-of term laboratory evaluations.** On a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree), mean ratings of the LIs’ performance characteristics were uniformly positive, ranging from 1.51 for approachability to 1.94 for clarity. The global rating (“The LI effectively facilitated my learning,” M = 1.86) was correlated strongly with each of the specific performance items, as shown in Table 2.

Mean ratings on the same scale for the laboratory experience ranged from 1.96 to 2.36. Finally, students rated the individual projects and overall impression of the lab on a final 5-point scale: 1 (very beneficial), 2 (quite beneficial), 3 (somewhat beneficial), 4 (minimally beneficial), and 5 (a waste of time). Means for the six projects ranged from 2.01 to 2.22. Table 3 shows the correlations of these items with rated overall impression of the laboratory.

**Student Performance**

**Pre–post critical thinking quizzes.** Two course instructors independently scored each scenario on a scale of 2 (correct identification of the critical methodological flaw), 1 (mention of some other legitimate flaw), or 0 (no response or an irrelevant one). Overall, interrater reliability for 1,320 paired judgments was .82. Alternate forms did not differ significantly in difficulty at pretest or posttest (both ps > .05). Students’ combined posttest scores (M = 8.98, SD = 3.03) were significantly higher than pretest scores (M = 7.19, SD = 3.27), t(210) = 9.15, p < .001, η² = .29.

**Postcourse critical thinking quizzes.** Two LIs independently scored the four scenarios on a 4-point scale: 0 to 2 points for correct identification of the critical methodological flaw and 0 to 2 points for identification of the optimal solution. Interrater reliability for 304 paired judgments was .87. The lifespan developmental psychology students who completed general psychology with a lab had a higher combined score (M = 9.18, SD = 2.81) than those who did not
Table 1. Pre–Post Questionnaire Data

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>Entry</th>
<th>Exit</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applying the scientific method to psychological issues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.87</td>
<td>3.57</td>
<td>0.70***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>2.55</td>
<td>3.70</td>
<td>1.14***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.26</td>
<td>3.12</td>
<td>−0.14</td>
</tr>
<tr>
<td>Writing formal laboratory reports (in APA style)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.17</td>
<td>3.86</td>
<td>1.69***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>1.96</td>
<td>3.90</td>
<td>1.94***</td>
</tr>
<tr>
<td>Interest</td>
<td>2.71</td>
<td>2.87</td>
<td>0.17</td>
</tr>
<tr>
<td>Coding observations in an unbiased manner</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>3.14</td>
<td>3.68</td>
<td>0.54***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>2.49</td>
<td>3.59</td>
<td>1.10***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.37</td>
<td>3.17</td>
<td>−0.20</td>
</tr>
<tr>
<td>Word processing skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>4.07</td>
<td>4.39</td>
<td>0.31***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>4.11</td>
<td>4.45</td>
<td>0.34***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.65</td>
<td>3.73</td>
<td>0.09</td>
</tr>
<tr>
<td>Statistical analysis skills using Minitab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>1.58</td>
<td>3.27</td>
<td>1.69***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>1.26</td>
<td>3.11</td>
<td>1.86***</td>
</tr>
<tr>
<td>Interest</td>
<td>2.48</td>
<td>2.61</td>
<td>0.13</td>
</tr>
<tr>
<td>Database searching using PsycINFO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.18</td>
<td>4.09</td>
<td>1.91***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>1.62</td>
<td>3.91</td>
<td>2.29***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.05</td>
<td>3.11</td>
<td>0.06</td>
</tr>
<tr>
<td>Understanding and evaluating psychological research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.87</td>
<td>3.69</td>
<td>0.82***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>2.34</td>
<td>3.61</td>
<td>1.28***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.78</td>
<td>3.37</td>
<td>−0.41***</td>
</tr>
<tr>
<td>Computer skills using psychology applications software</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.11</td>
<td>3.42</td>
<td>1.31***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>1.56</td>
<td>3.24</td>
<td>1.68***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.14</td>
<td>2.85</td>
<td>−0.29</td>
</tr>
<tr>
<td>Using computer-based laboratory equipment (e.g., iWorx)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.18</td>
<td>2.96</td>
<td>0.78***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>1.77</td>
<td>2.80</td>
<td>1.03***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.06</td>
<td>2.58</td>
<td>−0.48</td>
</tr>
<tr>
<td>Knowledge of scientific methodologies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.59</td>
<td>3.60</td>
<td>1.01***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>2.40</td>
<td>3.56</td>
<td>1.16***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.23</td>
<td>3.11</td>
<td>−0.12</td>
</tr>
<tr>
<td>Drawing conclusions from a psychological study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.97</td>
<td>3.79</td>
<td>0.81***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>2.49</td>
<td>3.74</td>
<td>1.25***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.68</td>
<td>3.47</td>
<td>−0.21*</td>
</tr>
<tr>
<td>Knowledge of psychology as a scientific discipline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comfort level</td>
<td>2.88</td>
<td>3.77</td>
<td>0.89***</td>
</tr>
<tr>
<td>Knowledge or experience</td>
<td>2.36</td>
<td>3.70</td>
<td>1.34***</td>
</tr>
<tr>
<td>Interest</td>
<td>3.52</td>
<td>3.45</td>
<td>−0.07</td>
</tr>
</tbody>
</table>

Note. N = 151. Scale is from 1 (very low) to 5 (very high).
*p < .05. **p < .01. ***p < .001.
experience a lab ($M = 7.46$, $SD = 2.46$), $t(73) = 2.85$, $p = .006$, $\eta^2 = .10$.

**Advanced course scientific writing.** Students in advanced psychology research courses who practiced writing APA-style manuscripts in our laboratory general psychology earned higher grades (out of 50 points) on second drafts of their first paper ($M = 46.18$, $SD = 2.73$) than those who completed the introductory course elsewhere ($M = 44.56$, $SD = 3.71$), $t(70) = 2.07$, $p = .04$, $\eta^2 = .06$. The instructor provided extensive feedback on all students’ ungraded first drafts. Nonetheless, the first group’s advantage on their second drafts was statistically reliable.

**Discussion**

According to standard definitions, psychology is the science of human behavior and mental processes. The subject of psychological inquiry is one that most students find fascinating. Yet, in too many cases, introductory psychology curricula gloss over the science aspect of the definition in favor of teaching “20,000 pieces of information about psychology” (Hailstorks, 2007, p. 17). This approach represents a missed opportunity and a departure from the origins of psychology born in the laboratory (Benjamin, 2000). Including laboratory experience in the general psychology course can help students more clearly appreciate and experience the ways psychologists search for answers to important questions.

We found that the laboratory curriculum described here has been successful in providing introductory psychology students meaningful experiences with scientific inquiry. Evaluations based on pre–post comparisons confirmed that students perceive themselves becoming more comfortable with and knowledgeable about the processes and methods of science. The results of critical thinking tests suggest that students became

Table 2. Laboratory Instructor Ratings and Intercorrelations

<table>
<thead>
<tr>
<th>Performance Characteristic</th>
<th>$M$</th>
<th>$SD$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequately explained purpose, method, goals of lab</td>
<td>1.94</td>
<td>0.99</td>
<td>.73</td>
</tr>
<tr>
<td>Organized and used time efficiently</td>
<td>1.91</td>
<td>0.93</td>
<td>.72</td>
</tr>
<tr>
<td>Knowledgeable and answered questions</td>
<td>1.86</td>
<td>0.96</td>
<td>.67</td>
</tr>
<tr>
<td>Approachable</td>
<td>1.51</td>
<td>0.88</td>
<td>.69</td>
</tr>
<tr>
<td>Available for help</td>
<td>1.77</td>
<td>0.92</td>
<td>.60</td>
</tr>
<tr>
<td>Graded fairly and consistently</td>
<td>1.84</td>
<td>0.97</td>
<td>.58</td>
</tr>
<tr>
<td>Provided constructive feedback</td>
<td>1.61</td>
<td>0.80</td>
<td>.68</td>
</tr>
<tr>
<td>Effectively facilitated students’ learning</td>
<td>1.86</td>
<td>1.03</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. $N = 198$. Scale is from 1 (strongly agree) to 5 (strongly disagree). Correlations are with ratings on the last item. For all $r$s, $p < .001$.

Table 3. Laboratory Ratings

<table>
<thead>
<tr>
<th>Item</th>
<th>$M$</th>
<th>$SD$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statements about the laboratory experience$^a$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercises were interesting ways to learn about scientific research</td>
<td>2.17</td>
<td>0.91</td>
<td>.55</td>
</tr>
<tr>
<td>Exercises were enjoyable change from lectures</td>
<td>2.36</td>
<td>1.10</td>
<td>.54</td>
</tr>
<tr>
<td>Laboratory reports effectively taught technical writing skills</td>
<td>1.97</td>
<td>0.90</td>
<td>.45</td>
</tr>
<tr>
<td>Laboratory reports required knowledge organization and understanding</td>
<td>1.96</td>
<td>0.84</td>
<td>.46</td>
</tr>
<tr>
<td>Exercises effectively introduced different research methodologies</td>
<td>1.93</td>
<td>0.80</td>
<td>.57</td>
</tr>
<tr>
<td>Ratings of individual laboratory exercises$^b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PsycINFO literature review on gender/culture topic</td>
<td>2.26</td>
<td>0.97</td>
<td>.50</td>
</tr>
<tr>
<td>Structured observation of children’s exploratory behavior</td>
<td>2.02</td>
<td>0.80</td>
<td>.53</td>
</tr>
<tr>
<td>Experiments on verbal memory phenomena</td>
<td>2.11</td>
<td>0.87</td>
<td>.54</td>
</tr>
<tr>
<td>Psychophysical correlates of stress</td>
<td>2.16</td>
<td>1.00</td>
<td>.60</td>
</tr>
<tr>
<td>Archival research on freshman survey data</td>
<td>2.22</td>
<td>1.06</td>
<td>.57</td>
</tr>
<tr>
<td>Overall impression of the laboratory component</td>
<td>2.10</td>
<td>0.89</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. $N = 198$. Correlations are with ratings on the last item. For all $r$s, $p < .001$.

$^a$Ratings on a 5-point scale ranging from 1 (strongly agree) to 5 (strongly disagree). $^b$Ratings on a 5-point scale ranging from 1 (very beneficial) to 5 (a waste of time).
more skilled at scientific reasoning. The benefit of this early engagement in psychological research persisted as students began a second psychology course, even when performance in the introductory course and time lag were controlled.

We believe that the introduction of the weekly laboratory experience has important ripple effects throughout our Psychology Department. All department faculty members, including those who do not regularly teach general psychology, actively collaborated on the design of introductory laboratory projects in their respective areas of expertise. They continue to consult on the ongoing revision of the laboratory curriculum. Student LIs made valuable contributions during the development and subsequent refinement of the projects. One cohort, on completing their tenure as LIs, reported a dramatic increase in their confidence; mastery of the subject matter; and improved oral communication, writing, and teaching skills (Faiola, Fernandez, Glad, McGray, & McNear, 2003). They became influential mentors to their fellow students. Their reflections echo the special advantages that accrue to student LIs (Hogan, Norcross, Cannon, & Karpiak, 2007; Newcomb & Bagwell, 1997). Finally, instructors of our advanced laboratory psychology courses have noted an advantage in the scientific writing and reasoning of students who had prior research experience in general psychology.

By using an inquiry-based approach, the revised general psychology laboratory fosters the enhancement of students’ scientific literacy, as well as their research, data analysis, critical thinking, and collaborative skills. These gains are relevant beyond the psychology course or major. They are useful for all students, regardless of their career objectives. Admittedly, we have not experienced a noticeable increase in the number of students declaring psychology as a major since we revised the laboratory component. On the other hand, we have not experienced a decline by intimidating prospective majors with a rigorous introduction to scientific psychology. The slight degree of eroded interest in computer applications might reflect the difficulties experienced by some students with malfunctions of the psychophysiological monitoring equipment. The realistic complexities of learning and doing science (research design, data analysis and interpretation, technical writing, and scientific thinking) might have also tempered some students’ early enthusiasm about the ease with which psychological science is conducted. Overall, however, the laboratory experience offers students a realistic preview of our department’s orientation to empirical research and helps to inform their choices in curriculum and career planning.

Nathanson, Paulhus, and Williams (2004) reported that science students who approach the introductory psychology course with a “scientific style of thinking” (p. 8) outperformed those who did not possess a science orientation. They argued that general psychology “may be attracting the ‘wrong’ types of students” (p. 8) and recommended tailoring college catalog descriptions of psychology to highlight its status as a science rather than an arts discipline, thereby attracting students with the “conscientiousness and tough-mindedness” (p. 8) of a scientist. However, their recommendation does not speak to the majority of students in psychology courses who are merely satisfying distribution requirements or prerequisites for another major. Our suggestion is to integrate research experience into the introductory course, thereby transforming all students’ scientific knowledge and reasoning in the process.

Although initiating a companion laboratory to the introductory psychology course involves a commitment of resources, it can be accomplished with judicious planning (Berthold et al., 2003). We would not be able to staff our laboratory without the assistance of student LIs. Given the benefits that accrue to students who serve in this leadership role and the advantages of providing meaningful work for student employment, departments can mount persuasive arguments for institutional support of LI appointments. We also recommend soliciting funds for physical space improvements and instructional technology from external agencies.

Psychology departments possess an excellent opportunity to introduce scientific psychology at a formative stage of students’ educational experience by integrating a research component into their first course. Laboratory projects promote active learning; hands-on experiences applying the scientific method; and practice interpreting, organizing, and writing the results of empirical studies. These transferable skills serve all students well.

We are sympathetic to Schwartz’s (2004) call for a problem-based learning curriculum for general psychology but question the efficacy of abandoning the lecture method altogether. We have found that augmenting classroom lectures, discussions, and demonstrations with well-integrated laboratory experiences provides fertile ground for engaging and challenging students at the outset of their journey as students of psychology.
References


Notes

1. We thank Randolph A. Smith and three anonymous reviewers for their helpful comments on earlier versions of this article.

2. This project was supported in part by National Science Foundation CCLI Grant 87926. Earlier versions of this article were presented at the January 2002 and 2003 National Institutes on the Teaching of Psychology, St. Petersburg Beach, FL, and at the Society for the Teaching of Psychology—American Psychological Society Teaching Institute, Atlanta, GA, May 2003.

3. Send correspondence to Thomas J. Thieman, Psychology Department, College of St. Catherine, St. Paul, MN 55105; e-mail: tjthieman@stkate.edu
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
A Video Introduction to Psychology: Enhancing Research Interest and Participation

Donald F. Sacco and Michael J. Bernstein
Miami University

To assess the extent to which a video about psychological research would heighten introductory psychology students’ interest and participation in research studies, we created a video about ongoing research at our university, the value of research participation, and course requirements for the research experience. Instructors in 4 courses (N = 471 students) verbally explained the research requirement and distributed a document detailing alternatives for completing it. In 2 of these courses, students also saw the video. Results indicated that students who viewed the video reported greater interest in research and completed more research participation hours than did students who did not view the video. We discuss the potential advantages of standardized, noncoercive procedures to encourage student participation in research.

Like most scientific endeavors, advancements in psychological knowledge are data driven. However, one of the foremost challenges for psychological research stems from the field’s unprecedented reliance on human participation; studying human behavior, from psychological disorders to social interactions, requires the study of actual people. To accommodate this demand for human participants in research, psychologists have historically relied on the use of human participant pools. For example, Jung (1969) evaluated 60 major universities and found that participant pools accounted for 90% of their behavioral science data. Corroborating these initial findings, a census of 366 psychology departments with graduate programs indicated that 74% of these facilities had participant pools, of which 93% reported recruiting participants from standard introductory psychology courses (Sieber & Saks, 1989).

Although such strategies for recruiting introductory psychology students for research participation help fulfill the demand for human participants, they are not without their drawbacks. One concern is that the methods used to encourage students to participate in psychological research may be coercive and potentially violate American Psychological Association (1992) ethical standards. That is, participants in these pools might consider the alternatives to actual research participation inequitable; they tend to be fewer in number, and participants perceive them as more effortful and time consuming (Sieber & Saks, 1989). Contrary to the concerns raised by existing recruitment strategies, recent evidence suggests that students themselves prefer participation in experiments to alternatives such as writing a paper. For example, self-reports have indicated that students consider research participation more interesting and more likely to facilitate learning than alternative research options (Trafimow, Madson, & Gwizdowski, 2006). These more recent findings are especially important when coupled with the fact that students report greater course satisfaction and educational benefits after volunteering for participation in actual research studies, compared to completing other research requirements (Bowman & Waite, 2003). In addition, Wade and Tavris (2003) suggested that participation in psychological experiments can help debunk students’ beliefs in myths regarding the nature of psychological research.
Because converging evidence indicates that research participation might be associated with educational benefits and course satisfaction, we considered techniques for encouraging such participation in a noncoercive, constructive manner. Reasoning that a video might encourage more participation than would an explanation on a syllabus or by an instructor, we created a brief video about psychological research. The video defined the field of psychology, described the kind of research conducted at the university, explained the value of students’ individual participation, and outlined the university’s specific procedures for research participation. We showed the video to students in two introductory psychology courses, whereas students in two other introductory psychology courses did not see the video (control condition). We hypothesized that students exposed to the video would report greater understanding of the research requirement, greater interest in psychological research, and greater interest in getting directly involved in psychological research than would students in the control condition. In addition, we hypothesized that students exposed to the video would ultimately complete more hours of their research requirement through direct participation in psychological studies (rather than through alternative options) than would students in the control condition.

Method

Participants

Introductory psychology students (N = 463, 171 men and 292 women) at a midsize Midwestern university volunteered to participate and received partial course credit for doing so. We randomly assigned each of four available classes, taught by different instructors, to either an experimental video condition (n = 177 students) or a control, no video condition (n = 286 students).1

Materials

The 8-min video included multiple components. First, it showed how psychologists use research to help solve real-world problems and to create a more complete understanding of human behavior. Second, it displayed a variety of departmental laboratory settings (e.g., clinical, social, and cognitive) and the equipment that students could expect to encounter in these laboratories. Third, it emphasized the rationale for the research experience. Fourth, it summarized ethical principles, such as confidentiality, that protect research participants. Fifth, it identified how students may complete the research requirement by writing papers or by participating in research. Finally, it detailed procedures for signing up for and participating in research, including informed consent and debriefing. This video (.mp4 format) is available to download from http://www.units.muohio.edu/psychology/ugrad.html at the link titled “Video: Psy 111 Brief Introduction to Research,” for viewing with QuickTime.2 We designed the video using Microsoft Windows Movie Maker (Version 5.1; 2001).3

Procedure

On the second day of class, in both the control and experimental conditions, professors outlined the ways to fulfill the research requirement and provided students with a document explaining the alternatives. This followed a standardized departmental protocol given to all students in the introductory classes. This included identical handouts explaining the requirements, why it is an important component of the course, as well as the ways in which students can complete this segment of the course. All introductory instructors are required to use and review this handout when explaining the research requirement to students. Along with the previous procedures, participants in the experimental condition then viewed the video. Next, all participants used Likert scales, ranging from 1 (not at all) to 7 (very much), to indicate the extent of their agreement with six items: interest in psychology, understanding of the research requirement, comfort with participating in research, interest in the kind of research conducted at their university, knowledge of what to expect when entering a psychological study, and interest in research involvement beyond the research requirement.

---

1The Miami University student body is predominantly White; students with other ethnic backgrounds comprised approximately 9.5%, of the Fall 2008 enrollment.

2Individuals can download the video by clicking on the link and choosing the “Save Target As” option from the drop-down menu. Furthermore, individuals may access the video from the YouTube Web site at http://www.youtube.com/watch?v=fqgLZQpdOi4

3Microsoft Windows Movie Maker is standard software on most Windows operating systems. Individuals can download this software at http://www.microsoft.com/downloads/search.aspx?displaylang=En
Table 1. Perceptions of and Participation in Psychological Research as a Function of Condition

<table>
<thead>
<tr>
<th>Statement</th>
<th>Control Condition</th>
<th>Video Condition</th>
<th>Statistic</th>
<th>Effect Size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>I am interested in psychology.</td>
<td>5.78</td>
<td>.95</td>
<td>5.79</td>
<td>.92</td>
</tr>
<tr>
<td>I understand the research requirement for</td>
<td>5.91</td>
<td>1.09</td>
<td>6.02</td>
<td>1.06</td>
</tr>
<tr>
<td>this class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am not comfortable with the idea of</td>
<td>2.65</td>
<td>1.60</td>
<td>2.57</td>
<td>1.60</td>
</tr>
<tr>
<td>having to be part of psychological research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am interested in the kind of research done</td>
<td>4.85</td>
<td>1.07</td>
<td>5.19</td>
<td>1.09</td>
</tr>
<tr>
<td>here at Miami.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I feel like I know what to expect when I</td>
<td>3.67</td>
<td>1.37</td>
<td>4.28</td>
<td>1.33</td>
</tr>
<tr>
<td>enter a psychological study.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think I would like to get involved in</td>
<td>3.68</td>
<td>1.51</td>
<td>3.99</td>
<td>1.41</td>
</tr>
<tr>
<td>research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total hours completed via direct participation.</td>
<td>11.31</td>
<td>2.52</td>
<td>11.89</td>
<td>1.56</td>
</tr>
</tbody>
</table>

Note. For the six questions, higher numbers indicate greater agreement with the statement.

All participants indicated their sex and year in school. Subsequently, to document behavioral outcomes, we obtained research participation reports for all participants every 2 weeks throughout the semester.

Results

Our primary interest was to document the extent to which an introduction to research video presented at the beginning of an introductory psychology course would influence students’ interest in, perceptions of, and participation in a required research experience. We conducted separate multivariate analyses on all of our dependent measures to examine the possible interaction between class instructor and condition as well as participant sex and condition. No significant interactions emerged across any of the measures (all ps > .05). Thus, we compared the responses of control participants (no video) to participants in our experimental condition (video). As shown in Table 1, independent samples t tests indicated that participants who viewed the video reported greater interest in the type of research conducted at the university, greater knowledge of what to expect when entering a psychological study, and greater interest in becoming involved with research than did participants in the control condition. Furthermore, our behavioral assessment of research participation indicated that of the 12 required hrs of research experience required for introductory courses, students in the experimental condition completed more hours of this requirement via direct research participation than did participants in the control condition.

Discussion

Recent research (e.g., Bowman & Waite, 2003; Rosell et al., 2005; Trafimow et al., 2006) has indicated that students in introductory psychology courses benefit more from direct participation in psychology experiments when attempting to meet their research experience requirement than from the alternatives to direct participation, such as writing papers. Based on such findings, our current goal was to develop a standardized method of familiarizing students with psychological research that would increase their enthusiasm in direct participation in a noncoercive manner. To facilitate this process, we developed a video to familiarize introductory students with the types of psychological research being conducted at their university, the process by which that research was carried out, how to participate in this research, and the value of such research participation.

To test the pedagogical utility of this video, some participants viewed this video at the beginning of the
semester of their introductory course, whereas others did not. All students received the standard departmental handout that outlined the research experience requirement, which was also included in all class syllabi. Results suggest that the video had a significant influence on students' perceptions of psychological research beyond the standard information included in the course syllabi and handout. Although the effect was small, exposure to this video appeared to increase the likelihood of students completing more of their research requirement by way of actual experimental participation. In addition, given the current concerns over the potentially coercive nature of research requirements, we see this video as a tool for informing students about the nature of research involvement prior to their actual participation, thus enhancing the informed consent process integral to ethical scientific research. Moreover, the use of such a video could heighten a more general interest in the science of psychology.

Although the results of this study are promising, this research is not without its limitations. Participant assignment to conditions was on a class-by-class basis such that two introductory psychology classes viewed the video and two did not. There are inherent problems with using intact groups, including the fact that we were unable to equate the classes along any number of possibly important variables (e.g., different teachers). Moreover, this research did not pinpoint which aspects of the video served to increase interest and research participation, nor whether the video's influence stemmed from its comprehensive coverage rather than the video format. Future research might examine these issues, perhaps comparing the effect of a video to the effect of an instructor's presentation of a standardized, comprehensive introduction to research. Future research might also explore whether motivation or prior experience moderates the video's effectiveness. Nonetheless, this research suggests that using a video to familiarize students with psychological research, including an introduction to the types of research conducted at their university, the value of research participation, and information about how to participate in research, can have pedagogical and practical benefits.

References


Note

Send correspondence to Donald F. Sacco, Psychology Department, 100 Psychology, 90 N. Patterson Avenue, Miami University, Oxford, OH 45056; e-mail: saccodf@muohio.edu.
Active Learning Within a Lecture: Assessing the Impact of Short, In-Class Writing Exercises

Adam Butler
Kayah-Bah Phillmann
Lona Smart
University of Northern Iowa

Students in 2 large sections of an introduction to psychology course responded in writing during class to questions regarding material recently presented in lecture. After writing, they shared and discussed their responses with others. The exercises motivated attendance, and students generally reacted positively to the technique. There was some evidence that completing the exercises facilitated learning, although the exercises did not appear to stimulate intellectual activity outside of class. The technique is an easy and effective way to enhance a lecture and stimulate active learning during class.

The traditional lecture format is a remarkably efficient method to present course material to large classes (Bonwell & Eison, 1991; Cashin, 1985). However, the effectiveness of the lecture is limited by several factors, including a lack of feedback about student learning, promotion of passive listening by students, poor suitability for teaching higher order thinking, and unreasonable demands on student attention (Cashin, 1985). Fortunately, many of these limitations can be ameliorated by intermingling active learning exercises within the lecture format (Bonwell, 1996). We describe one such exercise in this article and report the effects of the exercise on student learning and reactions.

The short, in-class writing exercise we developed is an amalgam of two widely used active learning exercises: minute papers (Angelo & Cross, 1993) and think–pair–share (Johnson & Johnson, 1999). Minute papers are a classroom-assessment technique in which students provide written responses to short, general questions such as, “What is the most important thing you learned in class?” (Angelo & Cross, 1993). Although minute papers are commonly used at the beginning or end of a class period, they can also be interspersed through the period to provide a different learning activity and break up a lecture (Bonwell, 1996).

Think–pair–share is a collaborative learning exercise in which students discuss a question in pairs and then share their ideas with the larger class. The advantages of think–pair–share over more traditional discussion methods are that more students are involved in the discussion, embarrassment is minimized because students share ideas among smaller groups, and students have the opportunity to meet other students (Bonwell, 1996; McKeachie, 1999). As with minute papers, think–pair–share can effectively be used to break up a lecture.

We named the exercise described in this article CARDS because students responded on index cards. After the instructor presented a concept in lecture, students responded to a question in writing on a 4 × 6 index card. Unlike the more general questions commonly posed for minute papers, we directed the card questions at specific psychological concepts (see Table 1). We used cards when we believed a concept was important or when we found interesting questions to ask. When students finished writing their responses, they exchanged cards with other students and discussed their answers in groups of two or three. During the card exchange and discussion, the instructor and teaching assistants circulated around the large lecture hall, discussing the question and possible answers with small groups of students. At the conclusion of the exercise, the instructor reported to the class interesting answers that students generated or the correct answer if there was one. The exercise took approximately 5 min to complete. Approximately once a week, students turned in their cards at the end of the class period, and we recorded the cards as complete or incomplete or evaluated and rated the cards before recording. The cards accounted for 25% of students’ final course grades. We rated the cards only to increase between-subject variability for the purpose of grading. For the assessment reported in this study, we scored the cards only as complete or incomplete. We predicted that students would have positive reactions to the card technique and that it would enhance learning.

Method

Participants

Students enrolled in two sections of introduction to psychology taught by the first author at a Midwestern state university participated in this study. The larger section (n = 125) met three mornings a week for 50-min periods, and the smaller section (n = 79) met one night a week for 2½ hr.

Procedure

Over the course of a semester, students responded to approximately 50 card questions. On 12 occasions, we pre-
sented a card question to one class but not the other and then linked these card questions to multiple-choice questions on the exam. Even when we did not present a card question to a section, we discussed the same material in class, and students received identical study guides for exams. To assess the effect of the differentially presented cards on learning, we compared the percentage of correct responses on the linked exam questions across the two sections. Table 1 presents a sample card question and corresponding exam question.

### Measures

#### Student Learning

We linked 12 card questions, 6 presented to each of the two sections, to questions on the exams. We measured learning by examining the percentage of students who answered the corresponding exam question correctly.

#### Student Reactions

We assessed student reactions with a questionnaire at the end of the semester and by compiling statements made about the exercise on the official university course evaluations. The questionnaire contained 5-point Likert scales measuring attendance motivation, perceived learning, exercise enjoyment, and intellectual stimulation. The anchors for the first two scales ranged from 1 (strongly disagree) to 5 (strongly agree), and the anchors for the last two scales ranged from 1 (never) to 5 (very often).

### Results

#### Student Learning

To assess student learning, we conducted binomial tests of the percentage of students answering each exam question correctly. We included only those students who turned in a card corresponding to the exam question. The test value was the percentage of students from the control group (i.e., the section that did not see the card question) who answered the exam question correctly. The results, presented in Table 2, show that 4 of the 12 tests were significant in the predicted direction. One of the 12 tests was significant in the opposite direction, and 7 tests showed no significant differences. A cursory inspection suggested that the exam questions showing a significant difference were quite similar to the card questions, as in Table 1, whereas the exam questions that did not show a significant difference tended to diverge more from the card questions to which they were linked.

#### Student Reactions

Descriptive statistics and scale intercorrelations for the reaction measures appear in Table 3. The mean rating for the attendance motivation scale was above the scale midpoint, indicating that the prospect of having to turn in a card was a motivator for attendance. This finding may have occurred
because the cards counted for 25% of the total points available in the class. The mean rating for the perceived learning scale was also high, indicating that students believed the cards facilitated their learning. This belief was also reflected in student comments on the course evaluations. Although the mean rating on the enjoyment scale was only at the scale midpoint, comments on the course evaluations indicated that some students found the cards interesting and stimulating. There was a significant negative correlation between attendance motivation and exercise enjoyment, indicating that students who attended class merely to turn in a card tended to enjoy the exercise less. It does not appear that the cards inspired students to seek information or discuss material outside of class, as the mean rating on the intellectual stimulation scale was relatively low.

Discussion

Our evaluation of CARDS indicates several positive effects on student reactions and provides some evidence that the exercise facilitated student learning. The exercise motivated students to attend class—an issue that is often a problem in large sections, although this effect may not have occurred if the cards did not contribute to the students' grades. Students also believed that the exercise was engaging and that it helped them learn the material. We did not find evidence that the cards stimulated intellectual activities outside of class. Perhaps the best indicator of student reactions is that over two semesters, three course sections, and approximately 400 students, no one ever made a negative comment about CARDS on the course evaluations.

We found that the CARDS improved exam performance on one third of the questions we tested. Although this finding is not overwhelming evidence that the exercises facilitated learning, it is important to note that both sections received the same course content as well as detailed study guides describing the content of the exams. Given that we did not return the cards to the students, it is unlikely that the improved exam performance was due merely to memorizing the card question and answer. Instead, the beneficial effect on learning may have occurred because the card questions provided an opportunity to practice recalling material and to receive immediate feedback about the answer.

A principal benefit of the card exercise is its flexibility for use within a variety of classroom contexts, for a variety of purposes (e.g., attendance checks), or for virtually any subject area in psychology. The exercise is easy to administer because it requires little time to prepare and evaluate and involves no technology beyond pencil and paper. Instructors of large classes would need to monitor the amount of noise generated by the discussion. However, given the benefits of CARDS, psychology instructors should react as favorably to the technique as their students do.

References


Notes

1. The first author developed the idea for this technique while he was a University of Wisconsin Teaching Fellow.
2. We thank Helen Harton for her comments on a previous draft of the article.
3. Send correspondence and requests for a complete list of card and exam questions or criteria for rating cards to Adam Butler, Department of Psychology, University of Northern Iowa, Cedar Falls, IA 50614–0505; e-mail: adam.butler@uni.edu.

Table 3. Scale Means, Standard Deviations, and Correlations for Exercise Evaluation Questions

<table>
<thead>
<tr>
<th>Scales</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance motivation</td>
<td>3.68</td>
<td>1.03</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived learning</td>
<td>3.45</td>
<td>0.89</td>
<td>–.11</td>
<td>.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise enjoyment</td>
<td>2.92</td>
<td>0.71</td>
<td>–23*</td>
<td>.49*</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Intellectual stimulation</td>
<td>1.55</td>
<td>0.64</td>
<td>–.04</td>
<td>.22*</td>
<td>.32*</td>
<td>.67</td>
</tr>
</tbody>
</table>

Note. Ratings were made on a 5-point scale ranging from 1 (strongly disagree/never) to 5 (strongly agree/always). Cronbach's alpha reliability coefficients are on the diagonal of the correlation matrix. *p < .01.
Sixty-five undergraduates participated in a small-group activity designed to help them apply the findings from classic studies of conformity, obedience, and social roles. Students designed and demonstrated a study to illustrate the influence of obedience, conformity, or social roles in a real-life context. Each group generated 3 variables and described how each variable might moderate the power of social influence in their situation. Evaluation indicated that most students found the activity enjoyable and helpful for understanding the role of social influence in their daily lives. Students preferred the activity to a lecture and recommended using the activity again in future introductory psychology classes.

As with most survey courses, introductory psychology presents an extensive amount of new material to students. Social psychology may span the length of two textbook chapters (e.g., Gray, 1991; Westen, 1999). Unfortunately, social psychology is often taught at the end of the semester (e.g., Gray, 1991; Kalat, 1996; Myers, 1996; Westen, 1999), which is a time when student interest and volition is relatively low (Schallert, Reed, Turner, & McCann, 1997). To increase student interest, I designed an activity that allows students to apply basic knowledge of three fundamental social psychological concepts: obedience, conformity, and social roles.

It is common for introductory psychology instructors to devote a day of lecture or film clips to the topic of social influence. Instructors may offer brief descriptions of Milgram’s (1963, 1974) obedience studies, Asch’s (1956) conformity studies, and Zimbardo’s (1973; Haney et al., 1973) prison study (Psychology Classics, Brothen & Spalding, 1992). Some instructors may ask students to engage in small skits or simulations (Wann, 1993). Role-playing is popular and can increase student understanding of social influence as well as other concepts such as neural information transmission and patient–therapist relationships (Balch, 1983; Hamilton & Knox, 1985; Wann, 1993). Such active engagement in course content is linked to successful learning and retention (McKeachie, 1994).

The basic concepts of obedience, conformity, and social roles are not difficult for students to master. It is likely more difficult, however, for students to generate and critically analyze novel examples of social influence in their daily lives. I designed an activity to augment traditional role-play activities. The activity requires that students take a basic knowledge of social influence and subsequently use it to design their own unique demonstration of social influence. Students acquire the knowledge necessary to understand the requirements of the activity and then design their own demonstration and present it to the rest of the class, in just 75 min of class time.

Method

Participants

Sixty-five introductory psychology students at the University of Texas at Austin participated. Most were first- and second-year students. Seventy-five percent were White, 10% Asian, 13% Hispanic, and 2% were of other ethnicity.

Procedure

I began class by showing a 5-min video clip that included simulations of Milgram’s (1963, 1974) obedience studies, Asch’s (1956) conformity studies, and Zimbardo’s (1973; Haney et al., 1973) prison study (Psychology Classics, Brothen & Spalding, 1992). I then divided the class into thirds, assigning each third to one topic: obedience, conformity, or social roles. I displayed a transparency that listed the three to four pages in the textbook that covered each topic, and I gave students 8 min to read the textbook section matching their topic. Then, I distributed a handout with the following instructions:

In today’s world, we cannot replicate some of these classic experiments. But, let’s just say we could and that we want to see if the effects replicate in a different context (e.g., instead of a teacher–learner situation in Milgram’s study, let’s say we want to study obedience using a different situation).

1. In your group, decide on a real-life situation in which you think your concept—obedience, conformity, or social roles—may have a huge influence on people’s behavior. Some examples are drunk driving, criminal behavior, cheating on exams, hazing, gang behavior, working at a company like Dell, participating on athletic teams (men vs. women, captains vs. members), etc.

2. Design a study to test whether your concept (conformity, obedience, or social roles) has an influence on people’s behavior. WORK TOGETHER. Make your study as simple as possible.
3. Act out your study for the class. For example, if you plan to do something like Zimbardo’s prison study, you might want to hold up a piece of paper that says “Day 1” and then act out what happened on Day 1; then hold up another piece of paper for “Day 2” and act out big events that you think may have occurred on Day 2, etc.

4. Then, on the board, list for the class at least three variables that will influence how important your concept is (e.g., Under what conditions are people more likely to fall into a social role? Under what conditions are people less likely to obey an authority figure?). Do not write much on the board—keep it brief and to the point. If you can bring these variables into your act, great, but it is not necessary.

5. Remember your task: Make sure the other students in the class know what they need to know about your topic.

Go! You have 25 min.

Students could divide their third of the class in whatever fashion they chose. Students were familiar with working in small groups in the class, and most students worked in a traditional group size of 4 or 5 people. One student, however, led a group of 11 people in a very successful demonstration of employee obedience.

Group Work

Although my students were accustomed to completing brief activities and generally seemed to enjoy them, they showed an impressive amount of interest in this activity. For the 25 min during which students worked, I made a conscious effort not to peek over shoulders and instead offered advice only if a group requested my assistance. When 25 min had elapsed, various groups volunteered to present. Skits were very brief, averaging 2 to 3 min in length. Due to time constraints, 6 of 10 groups demonstrated their study. Although the remaining 4 groups did not present at all, students turned their materials in to me at the end of class and received credit for participating. Two examples of group skits may illustrate students’ creativity and insight in completing the activity.

Skit 1

Students in one group performed a demonstration of social roles. While one student held up signs displaying to the audience the day of each event, the other two students acted as employees in an office building. On Day 1, the two employees, Mark and John, chatted about their weekends as they worked side by side. On Day 2, John was promoted to position of vice president of the company. On Day 3, John asked Mark to bring him a cup of coffee, and Mark did so begrudgingly. Mark began to tell John about an experience he had had the night before, but John interrupted him with a request for the morning paper. The skit skipped to Day 5, when Mark brought John his coffee and paper without complaining—it was a routine now. On Day 8, Mark handed John his morning coffee, and John handed Mark some of his clothes to be taken to the dry cleaner. On Day 10, John requested that Mark call him Mr. Stone instead of John. Mark cringed but complied. The skit ended with Mark saying, “Do you need me to work overtime this weekend, Mr. Stone?”

The students discussed three variables that may have altered the effect of social roles: (a) age, such that if Mark had been older than John he may have been more resistant to his new role as John’s helper; (b) personality, such that if Mark had been more assertive and confident he would have been more resistant to his new role; and (c) sex, such that if both employees had been women, the transformation from employee to employer may not have been so rapid because the women may have been concerned about destroying their friendship.

Skit 2

Students in one of the conformity groups played the roles of witnesses to a mugging. In Part 1 of the skit, several men and women were casually observing as a woman was mugged. In the mugging, a man ran into the woman and they tussled briefly over her purse. The mugger grabbed the purse as he pushed the woman to the floor, and then ran off. In Part 2 of the skit, the witnesses to the mugging were being questioned, as a group, by a police officer who had arrived at the scene. The police officer asked the witnesses what happened, and one male witness described the mugging in detail, but left out the part about the purse being stolen. One of the women spoke up and said, “Yeah, and he took her purse!” The witnesses debated for a short time—no one else saw the mugger take a purse. The woman defended her story for a little while and then folded, unsure of her own recollection of the event.

The students who designed the skit discussed three variables that may have impacted conformity: (a) individual versus group questioning, such that the woman would have been less likely to conform had she been questioned individually; (b) presence or absence of a partner, such that the woman would have been less likely to conform if just one other person had supported her story; and (c) relatedness to victim, such that if the woman had been related to the victim she may have been less likely to doubt what she saw.

After each skit was over, I led the class in a brief discussion. For example, after Skit 2, I asked whether the woman’s behavior illustrated normative or informational conformity. Then I asked students to describe how they would test the presenting group’s predictions. Thus, through these group demonstrations, students applied their knowledge of all topics, not just their own.

Evaluation and Discussion

Students evaluated the activity during the final 5 min of class. The results appear in Table 1.

On a scale ranging from 1 (strongly disagree) to 4 (neutral) to 7 (strongly agree), 78% agreed or strongly agreed that the activity was useful in accomplishing its objectives. Eighty percent disagreed or strongly disagreed with the statement, “The activity was a waste of time.” Sixty percent disagreed or strongly disagreed with the statement, “I would have preferred to have a
lecture on the information.” Finally, 74% agreed or strongly agreed that they would recommend using the activity again in other introductory psychology classes. These self-report data, however, do not directly assess student learning. Future studies of this technique could include two groups, one that completes this activity and another that receives a lecture, and then assess group differences in exam performance.

The evaluations are encouraging, for at least two reasons. First, 65 students completed the activity in one class session. Although 65 is not as large as some classes, small-group work and class presentations are often difficult to accomplish with more than 30 or 35 students. Second, students at large universities such as the University of Texas are accustomed to classes in which active learning is not a primary component of their education. Students often expect to be passive recipients of new information and are generally not required to apply the information they receive until they have an exam in hand. I find it encouraging that, in addition to completing the activity in a mature and creative manner, the majority of the students enjoyed the activity. The success I had with this activity illustrates that active learning strategies—specifically, small-group activities—can be effective even at large institutions and in relatively large classes. In larger classes and in 50-min sections, instructors might ask students to read the material before coming to class. Instructors might also choose to random two or three groups to present, thereby saving time and possibly increasing students’ perceived accountability for learning the material.

Instructors can implement this technique with other topics such as stages of cognitive development, classical conditioning, and positive and negative reinforcement—topics that can be relatively difficult for students to master. If the activity is brief and structured and students perceive it as necessary for thorough comprehension of the material, then students can be active participants in their education and thus be more effective learners, even in large classrooms.

References


Notes

1. I thank Traci Giuliano, JoyLynn Reed, and Wendy Domjan for their helpful suggestions on an earlier version of this article.

2. Send correspondence to April L. Bleske-Rechek, Vanderbilt University, Department of Psychology and Human Development, Box 512 Peabody Station, Nashville, TN 37203; e-mail: april.bleske@vanderbilt.edu.
ing periods (e.g., sustained coaching and practice of the skills required for effective teaching).

Although most respondents reported observation of GTA teaching and feedback, only about half reported using videotaping procedures. We encourage videotaping because it provides an accurate record of all aspects of GTA classroom performance. It is also a useful method to preserve GTAs’ classroom performances for future class discussion, feedback, and critique.

We discovered that respondents addressed certain important topics less frequently than other important issues (e.g., how to teach critical-thinking skills, institutional academic policies, diversity issues, social skills essential to effective teaching). Failure to cover such issues is unadvisable for two reasons. First, as some authors have argued (e.g., Nummedal & Halpern, 1995), teaching critical-thinking skills to undergraduates is essential to their development as independent and effective problem solvers. Second, GTAs who are unaware of institutional policies, insensitive to diversity issues, or deficient in the social skills necessary to establish rapport risk misinforming or offending their students.

We also urge more faculty who offer the teaching of psychology course to include content related to the use of electronic technologies. Such content may range from simple tasks, such as learning how to use an overhead projector, to more complex tasks, such as learning how to prepare and use a PowerPoint® presentation or developing Web-based resources. Learning to use electronic technologies may be useful in developing more effective, integrated, and thorough class presentations.

The teaching of psychology course must include content focusing both on the basic techniques of sound teaching and “people skills” to maximize chances that it will fully prepare GTAs for their first classroom teaching experiences (cf. Mueller et al., 1997). Indeed, such content, combined with exposure to the teaching literature, constructive feedback, and writing assignments, seems likely to maximize chances that the new psychology professorate will develop the skills and self-efficacy necessary to meet the challenges that await them (Prieto & Meyers, 1999).

References


Notes

1. We thank members of the EDGE group for research on the teaching of psychology at Auburn University for their criticisms on an earlier version of this article. We also thank Randolph A. Smith and three anonymous reviewers for their helpful comments on an earlier version this article.

2. Send correspondence to William Buskist, Psychology Department, Appalachian State University, Boone, NC 28608; e-mail: buskistw@appstate.edu.

Using Case Studies in Introductory Psychology

Julie A. Leonard
Kirsten L. Mitchell
Steven A. Meyers
Jacqueline D. Love
Roosevelt University

The case study is an active learning strategy that allows introductory psychology students to grapple with course content. To promote the successful development and use of case studies in introductory psychology courses, we present guidelines for implementing this technique as well as an illustration. Steps involved in using case studies include (a) determining the goals for the exercise, (b) selecting and narrowing a topic, (c) developing case material, (d) creating an inquiry component, and (e) assessing the effectiveness of the exercise.

The case study is one active learning strategy that allows introductory psychology students to grapple with course content. In psychology and other disciplines, its use correlates with positive learning outcomes such as increased student engagement, problem-solving abilities, and decision-making skills (e.g., Fisher & Kuther, 1997; Sudzina, 1999; Vernon & Blake, 1993). Moreover, analyzing case studies promotes critical thinking, broadens students’ perspectives on psychological issues, and fosters independent learning (McDade, 1995).

Case studies may be particularly well-suited for introductory psychology courses. Analyzing and discussing case mate-
Case studies can provide students with opportunities to evaluate data, identify important concepts, develop hypotheses, and create or defend arguments (McKeachie, 1999). These objectives contrast with goals best accomplished by lectures, such as acquiring factual information (Davis, 1993; McKeachie, 1999). Thus, the particular student competencies that instructors want to develop will shape the instructions for the activity as well as the case details.

After reflecting on their instructional objectives, psychology faculty must select a topic area. Instructors often use case studies to teach ethics (e.g., Costanzo & Handelsman, 1998) and abnormal psychology (e.g., Perkins, 1991). However, faculty can incorporate case studies when teaching many other subjects within introductory psychology, such as neuropsychology (Morris, 1991), child development (McManus, 1986a), or research methods (McBurney, 1995).

More important, instructors need to narrow the topic area to appropriately address it in the case study. Student feedback may provide instructors with valuable information about potential topics. Which areas typically prompt many student questions? What generates the most discussion or debate in class? Topics for case studies often address controversies (e.g., Ford, Grossman, & Jordan, 1997) and generally do not have one correct answer (McBurney, 1995).

To exemplify, our introductory psychology students often have difficulty objectively examining the topic of prejudice within the social psychology unit. Our goals for the case study exercise described subsequently include allowing students to (a) apply the course concepts of stereotypes, prejudice, and attribution to case material; (b) examine the interrelation between attitudes and behaviors; and (c) examine a societal implication of prejudiced attitudes. The focus of our case study is sufficiently broad to permit simultaneous analysis of several related topics (e.g., attribution, attitude formation, attitude change). However, it is appropriately narrow to permit in-depth exploration (i.e., we exclude other social psychological topics; e.g., social influence or cognitive dissonance).

Instructors need to construct the case study itself. Successful cases are accounts of realistic situations (Hoover, 1980). They can be actual or fictitious scenarios, vary in length, and focus on one or more problems. Instructors can develop cases themselves; find case material in scholarly publications, newspaper articles, films, or literature (e.g., Chisler, 1990; Logan, 1988); use those contained in ancillary materials for introductory psychology textbooks (e.g., Bolt, 1996); or encourage students to write their own cases (e.g., McManus, 1986b; Ortman, 1993). Cases can also include actual or simulated data to supplement narrative text (e.g., Witte, 1998).

Instructors should use a well-written story that contains important details so that students can engage substantively with the problem. It should be multilayered, with both obvious and subtle parts. The case study should be challenging and allow for student questioning. It should not, however, be overly complex such that students become confused when trying to follow the story line. Excessive intricacy detracts from the exercise and results in students feeling overwhelmed. Instructors need to be cognizant of students’ abilities and construct the case study to match this level. In our brief illustration, we present students with the following case:

You are a high school instructor who has been teaching for the past five years at a predominantly White, affluent high school. Because someone with more seniority has bumped you, you have been transferred to another school. It is located in the inner city, and virtually all its students are African Americans; it is on academic probation due to low achievement scores on standardized tests. You fear that some of your current students may be gang members because of the way that they are dressed. Frankly, you are concerned about your safety. Not only are you scared of being challenged by your students, but also of traveling to the neighborhood in which you now work. You have requested a transfer to a school in a “better neighborhood,” but the administration has told you that reassignment is unlikely before the school year is over.

Instructors need to add an inquiry section to allow students to apply their knowledge to the case. These questions should not only assess students’ ability to recall, comprehend, and apply relevant concepts, but questions should also focus on developing critical thinking skills such as analysis, synthesis, and evaluation. Moreover, questions should follow a logical sequence and should appear in an order of increasing difficulty. An appropriate number of questions proportional to the amount of time allocated for the exercise should follow the case study. Students may complete these questions individually, in small groups, or through class-wide discussion. Faculty may further assign written analysis of the case outside of class.

To illustrate the points described previously, we pose the following questions to students in our sample case study. Using Bloom’s (1956) taxonomy of educational objectives, we include a parenthetical note after each question to clarify what skill that question fosters.

- What are the stereotypes and prejudices that the teacher in this example holds? (comprehension)
- According to attribution theory, how might a person with these ideas and feelings categorize the behaviors of the students in the school? (comprehension and application)
- How might this teacher’s attitudes affect his or her behavior in this school setting? (application)
- What are some ways in which this teacher’s stereotypes and prejudices could be changed? (application)
• Should teachers with subtle prejudices be allowed to work in predominantly minority schools? Why or why not? (synthesis and evaluation)

The last step involves assessing the effectiveness of the case study. Instructors can periodically use classroom assessment techniques (Angelo & Cross, 1993) to gauge students’ reactions to these exercises or to determine whether using case studies accomplishes their instructional objectives for the class. We further recommend that instructors provide students with feedback and assign grades at the end of case study exercises. This process (a) communicates to students that case analyses are an important part of class, (b) allows students to progressively refine their inductive and deductive abilities, and (c) increases students’ motivation.

Our students provided us with written feedback to open-ended questions about our use of case studies. They indicated that case studies help them comprehend abstract course material by providing concrete illustrations and opportunities to verify their understanding of core concepts. Moreover, our students emphasized that analyzing case studies increases their involvement, interest, and attention in introductory psychology. Finally, our students frequently underscored how case studies are both a welcome contrast and complement to lecture and discussion methods. More specifically, they enjoyed applying and integrating course material using real-life situations. However, our students also believed that lectures help them understand important concepts and better prepare them to examine case material. They similarly reported that class-wide discussions after they analyze cases are useful to verify the accuracy and thoroughness of their conclusions. In sum, the use of case studies not only allows instructors to teach material in a novel manner, but it also can provide students with a meaningful and enjoyable learning experience.

References


In Search of Introductory Psychology’s Classic Core Vocabulary

Richard A. Griggs
Montserrat C. Mitchell
University of Florida

Given the finding that current introductory psychology textbooks do not share a substantive common core vocabulary, we examined 2 related questions. First, was there a substantial core a half century ago before introductory texts became so lengthy and encyclopedic? Second, is there a classic core vocabulary (terms in the core vocabularies of both contemporary texts and those from the 1950s)? We did not find a substantial common core vocabulary, we examined given the finding that current introductory psychology textbooks do not share a substantive common core vocabulary, we examined 2 related questions. First, was there a substantial core a half century ago before introductory texts became so lengthy and encyclopedic? Second, is there a classic core vocabulary (terms in the core vocabularies of both contemporary texts and those from the 1950s)? We did not find a substantial core vocabulary for the 1950s texts and conclude that it is highly unlikely one has ever existed. However, a classic core vocabulary of over 100 terms does exist, and we discuss the importance of covering these terms in the introductory course.

In the last decade, three studies (Landrum, 1993; Quereshi, 1993; Zechmeister & Zechmeister, 2000) at-
Focused Interactive Learning: A Tool for Active Class Discussion

Helen C. Harton
University of Northern Iowa

Deborah S. Richardson
Florida Atlantic University

Ricardo E. Barreras
Graduate Center
City University of New York

Matthew J. Rockloff
University of Nevada at Reno

Bibb Latané
Florida Atlantic University

Focused Interactive Learning (FIL) is a tool for teaching psychological concepts through student participation in a focused discussion with other class members. Students from 5 upper and lower level psychology courses participated in FIL exercises in which they answered several multiple-choice or opinion questions on their own and then systematically discussed each item for about 2 min with other students before giving a final answer. FIL increased student test performance, helped them get to know other students in the class, and had a small effect on students’ self-reported participation and interest in psychology.

Along with the call for more active learning in the classroom has come a request for more research on active learning techniques (Bonwell & Eison, 1991). In this article, we offer an evaluation of Focused Interactive Learning (FIL), an active learning technique based on principles of dynamic social impact theory (Harton, Green, Jackson, & Latané, 1998; Latané, 1996). Students answered multiple-choice or opinion questions, then discussed each item with others in the class for about 1 to 2 min each. After discussion, they answered the items again. The instructor then led a short discussion of the items.

Active learning exercises are enjoyable (Lawson, 1995) and motivating to students (Watson, Kessler, Kalla, Kam, & Ueki, 1996). They help students learn (Lawson, 1995) and increase their confidence with class material (Townsend, Moore, Tuck, & Wilton, 1998). Despite these benefits, lecture is the dominant method used in most college classes, particularly in large ones (Bonwell & Eison, 1991). Although many instructors admit the advantages of active learning, obstacles such as limited class time, increased preparation needs, lack of materials or resources, large classes, and perceived risk limit the use of such approaches (Bonwell & Eison, 1991). FIL offers a way around each of these roadblocks.
Classroom hours are limited; instructors often think they must cover a certain amount of information during the term and worry that nonlecture activities will interfere with meeting this goal. FIL, however, requires as little as 10 min and enhances lectures by allowing students to process the information they have heard and instructors to quickly assess student understanding (Johnson & Johnson, 1994). FIL may also serve to redirect student attention after passively listening to a lecture.

Another obstacle to active learning is the time involved in planning (Bonwell & Eison, 1991). FIL overcomes this barrier, as it requires minimal preparation—just the gathering of some items for discussion. Testbanks usually contain multiple-choice questions from which instructors can select FIL items, further simplifying the task. FIL also requires minimal resources—only paper, an overhead projector, or a chalkboard for presenting the discussion items.

FIL is ideal for large classes in which active learning exercises can be unwieldy and traditional discussions difficult (for other active learning ideas for large classes, see Benjamin, 1991). At best, in a large class only a few students can participate in a traditional discussion. However, because everyone discusses the items simultaneously in FIL, all students are able to participate—an important requirement of active learning (Mathie et al., 1993). Embarrassment is minimized, both because students are speaking before a much smaller group and because they are getting to know each other better (McKeachie, 1999).

Focused small group discussions such as FIL are also relatively risk free (Bonwell & Eison, 1991). Students enjoy discussions, and working in groups can increase performance (Borreson, 1990) and help students learn better (Kraft, 1985). Because they commit to an answer before discussion in FIL, they have a position to defend and a chance to “warm up” (Meyers & Jones, 1993). FIL can be directly tied to test items, so students readily see the relevance of the activity (Mathie et al., 1993). FIL sessions may also encourage students to read and prepare before class so they will be better able to review material with other students.

FIL gives students immediate feedback on how well they know the material reviewed in the FIL exercises. Feedback on homework problems and examinations typically comes to students several days after they have completed the assignments, by which time they may have forgotten much of the information. Immediate feedback is an important component of active learning (Mathie et al., 1993).

In this study, we evaluated the effects of FIL on student learning, motivation, and attitudes in five psychology classes of different levels. We also compared student evaluations of two types of FIL exercises.

Instructions for FIL

Present questions using an overhead projector or handouts. Have students mark their answers to each question, then ask them to discuss each question for about 2 min in small groups or with the people sitting to the left and right of them (this saves time in large classes in which moving into groups can be disruptive). Provide clear instructions, asking students to systematically and thoroughly discuss each alternative rather than simply argue about which answer is correct. After the 2-min discussion of each question, ask students to mark their answers again. A full-class discussion of the answers may follow the group discussions.

Complex, thought-provoking multiple-choice items culled from test banks are good FIL items (see first item in Table 1). One could also use multiple-choice opinion questions with no single correct answer (e.g., choosing a research method to study a particular problem; see second item in Table 1). A third question type has students give their opinions on 6-point scales (see third item in Table 1). A mix of factual questions accompanied by opinion questions on the same controversial topics (e.g., intelligence testing, violence in cartoons, government control of altered states of consciousness) is another alternative.

FIL items should be truly arguable and interesting; otherwise the “discussion” consists only of offering support for one’s opinion, with little opportunity for analysis or insight. Participation is more lively when students have already acquired some baseline knowledge of the topic through readings or lectures than when the exercise deals with material they have not yet covered.

Method

We evaluated FIL in five classes at Florida Atlantic University (FAU), ranging from 21 to 250 students and from introductory to advanced undergraduate levels. FAU’s student body is 61% women with an average undergraduate age of 26. Ninety-two percent of students are Florida residents; 66% are White, 13% are Black, 11% are Hispanic, and 10% are “other.” Five instructors, from a first-time graduate student to an award-winning full professor, taught the courses.

FIL Class Procedures

We conducted FIL sessions in a large general psychology class (n = 250) on every other book chapter. The instructor selected six multiple-choice questions from her extensive test file for each session (e.g., first item of Table 1). She selected items that were relatively difficult (an average of only 65% of students got the FIL items correct on previous tests) and complex (i.e., applied or conceptual questions with lengthy response alternatives) so the alternatives could be reasonably discussed as possible correct answers. Students first answered all items on their own to give them a starting point for the discussions. Then students discussed a randomly chosen half of the items with their classmates before answering all questions again. We asked students to discuss only half of the items so that we could compare performance on discussion items to nondiscussion items, but because this manipulation did not seem to be effective, we collapsed the two for later analyses.

Two small (n = 30 each) sections of general psychology incorporated weekly FIL exercises. Students completed 13 FIL sessions. Four sessions consisted of three to six multiple-choice, test-review questions, and nine sessions contained one multiple-choice question and one opinion question.
Students in an upper-level social psychology methods course (n = 21) participated in weekly FIL exercises in instructor-assigned three-person groups. Items included objective test-review items and multiple-choice questions for which there were no single correct answers (e.g., second item of Table 1).

Students in an upper-level undergraduate aggression seminar (n = 36) discussed controversial aggression-related issues (e.g., appropriateness of punishment techniques and effects of media on aggression) in small, instructor-assigned groups after marking their initial opinions on their answer sheets. In each of the four FIL sessions, students discussed one opinion item (e.g., third item of Table 1).

### Student Ratings

Seventy-eight students from the small general psychology classes (n = 10 and n = 28), the social psychology laboratory class (n = 10), and the undergraduate seminar in aggression (n = 30) completed an evaluation of the FIL exercises at the end of the semester (we did not obtain these ratings from the large general psychology class). After reporting their gender, ethnicity, hours per week studied for class, and class attendance, they completed eight 7-point scales ranging from 1 (not at all) to 7 (very much). See Table 2 for specific items. Open-ended questions asked students what they liked best about discussing items with their classmates, what types of questions they enjoyed discussing the most, and which exercise or topic they enjoyed most and least and why. The first author categorized these open-ended responses.

In the two smaller general psychology classes, students rated three of the test-review sessions and eight of the opinion sessions immediately after their discussions, indicating how much they enjoyed the FIL exercise and how much they believed they learned from the session on separate 5-point scales ranging from 1 (not at all or nothing) to 5 (very much).

### Results and Interpretation

We collected data evaluating three aspects of FIL: (a) the effects of FIL on learning, (b) the effects of FIL on student attitudes, and (c) student preferences for test-review versus opinion FIL sessions.

---

### Table 1. Sample Focused Interactive Learning Items of Each Type

<table>
<thead>
<tr>
<th>Multiple-choice test review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which of the following physiological responses is the result of arousal of the parasympathetic nervous system?</td>
</tr>
<tr>
<td>a. slowing of heart rate</td>
</tr>
<tr>
<td>b. stopping of digestion</td>
</tr>
<tr>
<td>c. enlarging of pupils</td>
</tr>
<tr>
<td>d. quickened breath</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple-choice without a single correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>You are interested in the relationship between self-esteem, mood, and helping. You did a study in which you measured college students’ self-esteem, then put them into a good mood by giving them a free doughnut or a bad mood by telling them that the study would only count for one credit, rather than the two promised. You then observed whether they helped another student (actually a confederate) on an anagram task. You want to replicate the study before you publish. If you could only choose one of the following, which would you be most interested in modifying for your replication?</td>
</tr>
<tr>
<td>a. using a different participant population (adults, children, etc.)</td>
</tr>
<tr>
<td>b. using a different measure of helping</td>
</tr>
<tr>
<td>c. doing the study in a different city or country</td>
</tr>
<tr>
<td>d. using a different manipulation for mood</td>
</tr>
<tr>
<td>e. doing the study at a different time (i.e., years from now)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many critics have argued that standard intelligence tests are “culturally biased,” causing minorities to perform poorly on average. Attempts at culture-fair tests (e.g., Raven Progressive Matrices) have increased the relative performance of minorities, but show much less capacity in predicting a child’s school performance. What kind of tests should schools adopt given this trade-off? (circle one)</td>
</tr>
<tr>
<td>−3</td>
</tr>
<tr>
<td>culture fair</td>
</tr>
</tbody>
</table>

### Table 2. Student Ratings on End-of-Semester Focused Interactive Learning Evaluations in Three Classes

<table>
<thead>
<tr>
<th>Items</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much did you enjoy the small group discussion exercises?</td>
<td>4.23</td>
<td>1.65</td>
</tr>
<tr>
<td>To what extent do you feel the small group discussions:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>helped you to get to know other students in the class?</td>
<td>5.17</td>
<td>1.61</td>
</tr>
<tr>
<td>helped you to learn about psychology?</td>
<td>4.06</td>
<td>1.74</td>
</tr>
<tr>
<td>caused you to participate in class more?</td>
<td>4.04</td>
<td>1.86</td>
</tr>
<tr>
<td>increased your interest in psychology?</td>
<td>4.03</td>
<td>1.78</td>
</tr>
<tr>
<td>increased your preparation for class?</td>
<td>3.23</td>
<td>1.64</td>
</tr>
<tr>
<td>helped you to make better grades on the tests?</td>
<td>3.24</td>
<td>1.76</td>
</tr>
<tr>
<td>caused you to study more?</td>
<td>2.47</td>
<td>1.57</td>
</tr>
</tbody>
</table>

Note. n = 78. Scales ranged from 1 (not at all) to 7 (very much). Labels were provided only for the endpoints of the scale. Means that share the same subscript are not significantly different from each other, p < .01.
Learning

To assess the effect of FIL on knowledge/learning, we compared the proportion correct for FIL items (discussion and nondiscussion items used in the FIL exercises) and non-FIL items (those that had not appeared in FIL exercises) for the 197 students who completed all three tests in the large general psychology class using a repeated-measures ANOVA with FIL or non-FIL items as the independent variable. We also conducted a second repeated-measures ANOVA with FIL or non-FIL chapters (not including the FIL items) as the independent variable.

Students performed significantly better on the FIL items than the non-FIL items (M = .69, SD = .15 vs. M = .61, SD = .12), F(1, 196) = 100.20, p < .001, η² = .34, even though the instructor purposefully chose FIL items to be more difficult on average than the non-FIL items.1 Perhaps more important, students performed better on the textbook chapters in which they had done FIL exercises than on those for which they had no discussions, even when we removed the FIL items from the analysis (M = .64, SD = .12 vs. M = .59, SD = .13), F(1, 196) = 88.63, p < .001, η² = .31.2

Student Attitudes

To evaluate the effect of FIL on student motivation and attitudes, we examined student evaluations of the technique at the end of the semester in the four smaller classes. We performed a repeated-measures ANOVA with evaluation type (the seven areas in which students evaluated FIL’s effects) as the independent variable, followed by simple contrasts to determine how the evaluations differed. We also compared ratings across classes using a MANOVA with class as the independent variable.

As revealed in Table 2, students reported moderate enjoyment of the small group discussion exercises. Their evaluation of FIL’s effects on their feelings and behaviors, however, was mixed. Students believed that FIL was more helpful to them in some contexts than in others, F(6, 456) = 40.73, p < .001, η² = .35. They believed that FIL helped them to get to know other students in the class. They also rated FIL above the midpoint of the scale on its effects on their learning, participation, and interest in psychology. These ratings were not incredibly high, but they do offer a moderately positive appraisal of FIL. Students did not believe that FIL increased their class preparation or grades very much and believed it had even less of an effect on how much they studied for the course. According to the student raters, FIL had the most effect on interpersonal aspects of their class experience, followed by motivational aspects. Students did not believe that FIL affected their class-related behavior, although the objective test data presented previously suggest that FIL may have improved their grades, even if they did not realize it.

Students across all classes generally rated FIL equally. There was a marginally significant multivariate effect for class on the variables in Table 2, F(24, 194) = 1.56, p = .05, η² = .16. caused by a univariate effect for helping them make better grades on tests. Students in the aggression class attributed less grade improvement to FIL (M = 2.73, SD = 1.72) than did students in the other three classes (M = 3.56, SD = 1.72), F(3, 73) = 3.53, p = .02, η² = .13. In the aggression class, students discussed only opinion questions, which they may have perceived as less related to exam content.

Two open-ended items—“What did you like best about discussing items with your classmates?” and “What other comments, good or bad, do you have about the small group discussions?”—asked for student impressions of FIL. They reported several advantages to FIL, including learning about others’ points of view (mentioned by 53% of students), getting to know other students (22%), learning the material better (10%), and rethinking their own opinions (5%). Only 4 of the 78 students wrote that they did not enjoy the FIL discussions or find them useful.

Test-Review Versus Opinion Items

To assess the differential effectiveness of FIL sessions that consisted of test-review questions versus sessions that focused on opinions, we compared evaluations of the two types of items in the smaller general psychology classes with two mixed ANOVAs with item type and class as the independent variables. Students in one of the classes reported that they enjoyed and learned more from the test-review sessions than from the opinion ones, Class × Item Type interactions: enjoyment, F(1, 53) = 12.37, p = .001, η² = .19; learning, F(1, 53) = 9.10, p = .004, η² = .15; simple main effects: enjoyment, F(1, 27) = 29.31, p < .001, η² = .52; learning, F(1, 27) = 46.39, p < .001, η² = .63 (see Table 3), whereas in the other class students rated their enjoyment and learning on the two types of exercises equally. The differences between the classes may have been due to differences in teaching style of the instructors. In the class in which students preferred test-review items, the instructor stayed very close to the order and presentation of information in the textbook during lectures. Because students in this class were accustomed to staying very fact focused in class, they may have preferred FIL items that were consistent with that emphasis (i.e., test-review items). In the other section of the course, the instructor did not follow the textbook so carefully and sometimes discussed material that was not covered specifically in the textbook. Those students may, then, have been more accustomed to some departure from textbook material in the form of opinion discussions.

Three open-ended items on the end-of-semester evaluation that the smaller general psychology classes and the two upper-level classes completed asked students which types of questions they liked most and least. Of those who had a preference (68 of the 78 students), 73% preferred the opinion.
questions to test-review ones. There were almost as many particular topics mentioned as favorites as there were students, although students agreed that controversial topics elicited the most successful discussions.

Although one of the two small general psychology classes rated the multiple-choice questions more highly on the individual exercises, students stated that they preferred opinion questions on the end-of-semester evaluation. It is likely that opinion items were both most and least preferred. Getting put into a “group of slackers” (in one student’s words) or a particularly emotional or close-minded group may have made students’ discussion experiences for a particular question less enjoyable or rewarding, even though they enjoyed discussing opinion issues when the environment was more conducive. Supportive of this interpretation, 6% of students volunteered on the end-of-semester evaluations that their enjoyment of FIL was dependent on their group.

Conclusions

Students performed better on test items that were part of FIL exercises than on those that were not, and the effects of discussion seemed to generalize, as students also did better on their tests on chapters in which they completed FIL exercises than on those in which they did not. Students reported that the exercises helped them get to know others in the class and positively evaluations in all regards, however. Students did not be-

as well as benefit from, the activity. Instructors can easily in- incorporate FIL discussions into any psychology class (or class in any discipline, for that matter). FIL may be an effective supplement to lectures that gets students involved and interested and enhances their learning.

References


The Undergraduate Research Assistantship: An Analysis of the Benefits

R. Eric Landrum
Lisa R. Nelsen
Boise State University

This study documents and quantifies the benefits of serving as an undergraduate research assistant based on the results of a national survey of undergraduate psychology educators (N = 211). The survey consisted of a list of 40 potential benefits, skills, and abilities. Respondents rated each of the items on (a) whether their research assistants attain the benefit, skill, or ability and (b) the importance of each item to an undergraduate education in psychology. Factor analysis revealed 2 major themes: The first factor contained items relating to technical skills, such as math, statistics, writing, and effective communication, whereas items in the second factor pertained to interpersonal benefits. This study provides important information for evaluating the value of the assistantship experience.

Psychology has become an increasingly popular field of study among undergraduates, as the number of students majoring in psychology has risen dramatically in recent years. For 1996/1997, the number of baccalaureates awarded in psychology was 74,191 (National Center for Education Statistics, 1999). Students can therefore expect their career paths, such as graduate school or employment, to be highly competitive and should strive to develop skills and abilities that distinguish them from other psychology majors.

There are several opportunities for students to enhance an undergraduate education in psychology, and for students to make the most of their education, they should become involved in activities outside of traditional coursework (Landrum, Davis, & Landrum, 2000). One such opportunity is for students and faculty to collaborate on research or book projects. Jones and Draheim (1994) found their 8-month writing collaboration to be a highly beneficial experience for both student and teacher. Some of the student responsibilities included conducting library literature searches and critiquing drafts of a college textbook written by the professor. Their experience led to benefits for the student, such as improved writing and library skills, and the instructor got much-needed help with a time-consuming project.

The collaborative effort just described is much like the more commonly used undergraduate research assistantship. Educators seem to agree that research assistantships are beneficial to students. Evidence of the popularity of assistantships is revealed by examining the increasing number of papers presented at conferences by students (Kierniesky, 1984; Palladino, Carsrud, Hulicka, & Benjamin, 1982) and the increasing number of colleges utilizing the research assistantship as part of their undergraduate programs (Starke, 1985). Perlman and McCann (1999) found that, in 1996/1997 college catalogs, 28% of the institutions listed Research Participation as an available course. However, as Worthen (1969) noted, psychology educators tend to make assumptions regarding the innate value of the research assistant experience. In other words, educators may assume that one assistantship experience is just like another and that the experience is "inherently valuable" (p. 3). Faculty use a variety of models when working with undergraduate assistants, which leads to a wide array of skills, abilities, and benefits students can gain. Because the research assistantship is becoming increasingly popular, it is important to know what, specifically, about the experience is beneficial and how educators can maximize the benefits.

Limited research exists along these lines. However, studies to date have examined this question in two ways. Some examined the benefits as perceived by faculty members, and others examined the benefits as perceived by students. Faculty respondents reported that the experience is an important preparation for graduate school and that the experience generates excitement and curiosity about the research process (Kierniesky, 1984; Palladino et al., 1982). Additional benefits, such as improved critical- and independent-thinking...
A Motivating Exercise for the Introductory Class (and Beyond)

Louise Katz
Columbia State Community College

This article describes an exercise used with 12 community college psychology classes that produces high levels of motivation and offers the opportunity for a personally meaningful experience. Students choose a topic of personal interest related to psychology and investigate it through student-conceived field experiences and literature research. The exercise includes a poster presentation, handout, and oral presentation or, alternatively, a term paper and journal. At least 78 of 181 students reported selecting a topic they hoped would be useful in their personal life, and 88 reported that it had been. Many reported being helped with career decisions. In classes with oral presentations, students learned from each others’ research, and 93 reported finding listening to classmates’ presentations worthwhile.

Instructors often struggle to develop strategies to motivate students. In this article I describe an exercise that I have found very motivating with community college students, that can be carried out even with limited academic resources and may be used or modified for use with a variety of psychology courses in 2- and 4-year institutions. The exercise provides an academically and personally meaningful experience, greatly engages student interest in psychological topics, and creates appreciation of the value of psychological research through “hands-on” real-life experience. Brems (1994) reported that early familiarity with the research process may facilitate lessened anxiety in upper level courses. Others have discussed the value of journal writing (Connor-Greene, 2000), poster presentations (Baird, 1991; Chute & Bank, 1983), and oral presentations (Saunders, 1985) and have explored the value of out-of-class and typically preselected experiences for undergraduates (Clements, 1995; Ferraro, 1990; McCluskey-Fawcett & Green, 1992; Perry, Huss, McAuliff, & Galas, 1996; Sugar & Livosky, 1988). However, there is an absence in the literature of projects for freshmen where students both select their topic to investigate and fashion a means to explore it in the “real world.”

Method

At a rural community college, 199 predominantly freshman students in 10 introductory psychology classes and 2 life span development survey classes participated in the exercise. Students in 5 classes presented their work by submission of a journal and term paper, and students in 7 classes presented their work with oral presentation to the class; I discuss these two approaches separately. Approximately 73% of the students were women and 27% were men; some were returning adult students. Students chose a topic of personal interest related to psychology and prepared a one-page proposal, due 3 weeks after the beginning of the semester, that included a minimum of two proposed field experiences conceived by the student to investigate the topic and the proposed area or areas of reading related to the topic. Prior to the proposal due date, I discussed types of possible field experiences with the class, such as observations, interviews, attendance at meetings or support groups, surveys, and simple experiential projects, and I explained interview techniques, confidentiality, and securing of appropriate permissions. Also prior to the proposal due date, the class attended a presentation by the librarian on methods of library and database research, and students shared potential projects in a class discussion and received feedback from me. I returned written feedback on the proposals or, with some classes, met individually with each student during a class session to give oral feedback. Review of proposals gives the instructor the chance to ensure that projects do not pose ethical, student welfare, or institutional liability concerns. I reviewed proposals I had concerns about with another licensed psychologist knowledgeable in ethical matters, and I redirected these students to a related emphasis in their field experience, typically an interview. A few weeks later, I invited students to raise questions in class and offered them the opportunity to discuss their progress with me individually. I required five research sources for the exercise; I permitted two of these to be Web sites. The project as a whole represented 33% of the final course grade.

Presentation Approach

The presentation approach, used in seven classes, required students to present their work with an oral presentation, poster, and handout. Each student created a poster for a poster session held during the class session before oral presentations began. I explained possible approaches to poster organization in class and encouraged creativity. I advised students to include in their posters information gained from both their literature research and field experiences. Oral presentations took place over a period of one or more weeks near the end of the semester. Some students expressed anxiety about presenting a speech. It was helpful to discuss how to prepare for and present a talk. I made available a pamphlet on public speaking anxiety (Katz, 2000). Each student prepared a handout that covered the main points learned from literature research plus a list of references used and furnished a copy to each classmate and the instructor at the time of the oral presentation. I typically required a presentation of 20 to 25 min, followed by the opportunity for questions and comments. I found that a length requirement of at least 15 min encouraged students to prepare thoroughly. To permit alternatives for anxious students, I gave...
the options of videotaping the presentation and showing the videotape in class or making an appointment with me to discuss submitting their work in writing. I gave each student a grading sheet that gave feedback on each component of the presentation approach section. I assigned plus and minus letter grades on students’ projects as a whole, which translated to a set number of points toward the final grade.

**Journal and Term Paper Approach**

The journal and term paper approach, used in five classes, required students to present their work with a written journal and term paper due near the end of the semester. Each student prepared a journal that consisted of entries about thoughts and feelings prior to the field experiences and any changes in thoughts and feelings as a result of participation in the field experiences. I did not specify a length but would suggest doing so to guide students in writing appropriately detailed entries. Each student completed a term paper, minimum eight pages in length, based on his or her reading. Although I did not include a poster session or handouts when I used this approach, either or both could be included to provide a way for students to learn from classmates’ projects using minimal in-class time. I assigned letter grades as described in the presentation approach section.

**Results and Evaluation**

Eleven of the 12 participating classes (N = 181) formally evaluated the term project at the end of the semester. In an anonymous checklist evaluation, the majority of students (89%) evaluated the assignment as worthwhile; 93% indicated they had worked “very hard” or “pretty hard”; only 8% indicated “I just did what I thought I needed to in order to get by.” Thirty-two percent of students reported spending over 15 hr on their project from start to finish, and an additional 35% reported spending between 10 and 15 hr. In contrast, a minimum eight pages in length, based on his or her reading. Although I did not include a poster session or handouts when I used this approach, either or both could be included to provide a way for students to learn from classmates’ projects using minimal in-class time. I assigned letter grades as described in the presentation approach section.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special needs children</td>
<td>I chose this topic I guess in a way to help myself and it has done that. I feel like I am a better parent, I am more informed and have found new people to talk with about problems and joys. Thank you so much for giving me this opportunity to do this paper ... it has helped me so much.</td>
</tr>
<tr>
<td>Psychological effects</td>
<td>When I first started [this project] I felt I would never be able to work with very ill newborn babies. I thought maybe I would specialize in cardiac nursing. Now I think I would prefer to work in a neonatal ICU.</td>
</tr>
<tr>
<td>on nurses in ICU</td>
<td>This term project has opened my eyes. I was not aware that the effects could be so devastating. Looking back, I realize how close I actually was to danger.</td>
</tr>
<tr>
<td>Eating disorders</td>
<td>I found out a lot of interesting things, not only for my paper, but for my own personal knowledge. I finally understand why my parents did not stay together.</td>
</tr>
<tr>
<td>Aspects of love</td>
<td>I got all kinds of information. ... Their story is sooo interesting I could have sat there all day and listened. ... I went into this project not knowing much and feeling sorry for people with brain damage. I now feel respect. ... I no longer feel nervous. I kinda want to find more people and hear their stories.</td>
</tr>
<tr>
<td>Brain damage</td>
<td>The inside look at [schizophrenia] really opened my eyes. I will forever have a changed viewpoint. ... Overall, I was surprised how much I learned doing this project.</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>My beliefs have definitely been reinforced on the horrendous effects of alcohol. This research will stand as a lighthouse in my memory, helping me light the way for others by educating them to its dangers.</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>The whole experience has changed my outlook about a lot of things. I now volunteer every week at the [local] animal shelter and I’ve also adopted a Golden Retriever puppy from the shelter to begin training as a therapy dog.</td>
</tr>
<tr>
<td>Animal-assisted therapy</td>
<td>I feel that this project has confirmed and reinforced my decision to become a nurse.</td>
</tr>
</tbody>
</table>

Some findings of particular interest are that 78% picked a topic they hoped would be useful in some way in their personal life, and an additional 18% thought they might have done so. Also of interest, 88% indicated their project had, in fact, ended up being useful in some way in their personal life; 22% indicated they had made or were thinking about making one or more changes in their lives because of it. Students who completed oral presentations indicated a strong preference for oral presentations (70%) over a term paper and journal and overwhelmingly indicated (93%) that they found listening to the oral presentations of classmates worthwhile; 46% found listening "extremely worthwhile," and only 7% found listening "mostly pretty boring." Overall, students reported learning approximately equally from their field experiences and their literature research. Many students completed more than the required two field experiences. Patterns of evaluation response were very similar for the presentation approach (n = 85) and the journal and term paper approach (n = 96). Most students appeared to select as their topic a personal or career-related question, which they framed and investigated from a psychological point of view. Some reported being helped with career decisions or being offered jobs by organizations where they engaged in a field experience. For their field experiences, my students proposed naturalistic observations, interviews, attendance at meetings, surveys, and simple experiential projects. No students proposed experiments or correlational studies. If students had proposed these, it is unlikely they would have been approved, as our community college does not have a formal institutional review board, and faculty typically teach five large classes and time to supervise such projects would be scarce. Examples of students’ topics and comments appear in Table 1.
Conclusions

Experience with this exercise demonstrates that students, including those who may not be motivated to study for examinations, will find the combination of designing a research topic of personal significance and fashioning the approach to investigating it and a hands-on, real-life experience highly motivating. The motivational value of the exercise is apparent by the amount of time students reported spending on it compared to the amount of time they reported spending preparing for examination, by students’ positive written and oral feedback to me, and by comments repeated to me by colleagues. I believe the unique motivating element in the exercise is the students’ opportunity, under the aegis of the assignment, to uniquely shape their field experiences to satisfy a deeply felt need to know. Students did not select “easy” field experiences but fashioned personally meaningful ones. For example, a student whose career goal was pediatric nurse practitioner, who believed she had not yet confronted her feelings about coping with grief reactions in her chosen field, interviewed two nurses about how they coped with grief.

Initially, I used the journal and term paper approach. Later, I developed the presentation approach to enable students to learn from each others’ research, to alleviate substantial grading time due to class size, and to mitigate concerns about plagiarism. Knowing they will present their work to their peers seems to motivate students to work harder. However, it is difficult to find time to schedule oral presentations in classes with over 25 students, and students may sometimes misstate facts in oral reports. A strength of the project is it permits students who may do poorly on exams a chance to excel. The exercise with written approach takes minimal in-class time but, like any written assignment, is limited by the time of the faculty member. Although the assignment is challenging, and perhaps because it is challenging in a way students are eager to be challenged, I have found the level of students’ work on this project to be far superior to the level of students’ work in past years on other written and oral assignments, and I found many presentations quite provocative. The exercise offers valuable experience and makes the methods, possibilities, and value of research come alive.

References


Note

Send correspondence to Louise Katz, Department of Psychology, Columbia State Community College, P.O. Box 1315, Columbia, TN 38402–1315; e-mail: katz@columbiastate.edu.

Curriculum Review Using a Knowledge, Skills, and Abilities-Based Assessment of Alumni

Carrie B. Fried
John C. Johanson
Winona State University

The recent emphasis on assessment and career preparation has placed a spotlight on the skills that the undergraduate psychology major provides. Although there has been some assessment of the skills psychology graduates value and possess, there is little to no evidence linking the development of specific knowledge, skills, and abilities (KSAs) to specific psychology courses. We report the construction and results of an alumni survey designed to determine where students acquired specific KSAs. The results identified the psychology courses that teach specific KSAs. This assessment method may be useful for other departments as a guide for detecting weaknesses in their curricula, assessing the value of specific courses, and helping determine the best composition of the undergraduate psychology curriculum.

Questions regarding the usefulness of a psychology baccalaureate degree have existed for decades, but the recent emphasis on assessment has brought these concerns to the forefront. Several surveys have revealed that recent psychology alumni generally do not believe their degrees have prepared them well for careers (Borden & Rajekici, 2000; Finney, Snell, & Sebby, 1989; Grocer & Kohout, 1997; Nelson & Johnson, 1997). These surveys highlighted the need to better integrate the development of career relevant knowledge, skills, and abilities (KSAs) into a traditional liberal arts education in psychology. Students and teachers can easily find publications that list KSAs that psychology students possess (Appleby, 1997; Kruger & Zechmeister, 2001), KSAs that graduates find most useful on the job (Grocer & Kohout, 1997), and KSAs that employers value most (Appleby, 2000; Landrum, Davis, & Landrum, 2000). What seems to be missing, however, is a means to assess where students acquired those KSAs.
Students Teaching Students: An Experiential Learning Opportunity for Large Introductory Psychology Classes in Collaboration With Local Elementary Schools

Gary M. Muir and Gretchen J. van der Linden
St. Olaf College

Students in large, lecture-based introductory psychology classes often do not have the benefit of experiential learning (EL) opportunities due to logistical constraints. To overcome this obstacle, we developed an EL project in which introductory psychology students in small groups present some aspect of the course material to local elementary school classes. The project challenges undergraduate students to demonstrate a deep level of understanding of the presentation material. Such depth of understanding enables them to flexibly communicate material in an age-appropriate manner to kindergarten through 5th-grade students. Feedback results from undergraduate students, elementary school students, and teachers demonstrated that this project supported learning outcomes in undergraduates in a positive and enjoyable way for both student groups.

A particular type of EL opportunity, “supplemental student teaching,” effectively addresses both logistical teaching constraints and the need for a solid knowledge base in a subject (Nortcliffe, 2005). In addition, supplemental student teaching might be especially effective because of the public role it requires student presenters to take (Bargh & Schul, 1980). As Harel and Papert (1991 observed, “[learning] happens especially felicitously in a context where the learner is consciously engaged in constructing a public entity, whether it’s a sand castle on the beach or a theory of the universe” (p. 1).

There is also a clear service learning component to an EL project that has undergraduate students teaching in elementary schools. Service learning has been shown to promote a number of positive learning outcomes, including enhanced academic performance in undergraduates (Lundy, 2007) and increased understanding of course content (Hardy & Schae, 2000; Kretchmar, 2001). A further advantage of the supplemental student teaching in public model is that it can expose elementary school students to an area of science that they otherwise might not usually encounter until college. Psychologists are in favor of such a change; participants in the American Psychological Association’s 2002 Education Leadership Conference agreed that the teaching of psychology needs to be integrated into the curricula of primary and secondary schools (B. Murray, 2002).

We describe a supplementary student teaching in public model for undergraduate psychology students in large classes, such as the introductory psychology
classes at St. Olaf College, which regularly enroll more than 100 students. This “students teaching students” model provides large numbers of students with an EL opportunity as they develop and present a lesson in psychology to elementary school students. The advantage of this model over previously reported supplementary student teaching models (Elmendorf, 2006; Pezdek, 2002) is that it can be employed in large classes and requires students to spend significantly less time in the elementary school classroom. The project’s primary objectives were to: (a) provide college and university students in large classes with an EL opportunity and the chance to develop a deeper understanding of an aspect of psychology, (b) provide students with an opportunity to connect developmental psychology concepts with real children, (c) expose elementary students to the field of psychology, and (d) strengthen connections between the college campus and the wider community.

The Project and Participants

To establish the project, we conducted individual meetings to explain the goals and procedure of the project with three local elementary school principals, who subsequently agreed as a group to pilot the project across all three schools. Students in the introductory psychology course at St. Olaf College (a 4-year, liberal arts college) participated in the semester-long project over seven semesters (N = 774). Approximately 25 to 30 groups of 3 to 5 undergraduate students prepared presentations each semester for the three elementary schools. Following several group meetings, students chose presentation content from one chapter of the introductory psychology textbook and provided short descriptions of their proposed presentation to the instructor and two teaching assistants (TAs), who screened the descriptions for age-appropriateness, edited them, and arranged them into three sign-up sheets containing approximately 10 presentation topics, one for each of the three elementary schools. Once kindergarten through fifth-grade teachers had signed up for a presentation topic(s), the completed sign-up sheets were collected and students were notified of their presentation teacher, grade level, and teacher’s contact details. This provided students with at least 2 to 3 weeks to contact the teacher to arrange a mutually convenient time to present within a specified 2-week period at the end of the semester.

Creating an Age-Appropriate Presentation

One of the biggest challenges that students faced in this project was how to create an age-appropriate presentation. Because students did not know the exact grade level they were presenting to until approximately 2 to 3 weeks prior to the presentation, it was important to make sure that groups developed an outline that was flexible enough to be adapted to the final age level. To assist with this process, approximately 3 weeks prior to the presentations, we covered the developmental psychology part of the course in lectures, and a St. Olaf College education professor guest lectured on creating age-appropriate presentations for elementary school-age children.

Examples of Presentation Content

The undergraduate students showed flexibility and creativity when faced with the task of adapting content for various age groups. For example, one group presenting on sleep came dressed in pajamas and pretended to wake up one member of the group while he slept to demonstrate how dreams are different depending on the part of the sleep cycle in which one is awakened. Another group engaged elementary students by helping them act out the modal model of memory; some children played the roles of sensory memory, working memory, and long-term memory, and other students served as the control processes, running between memory stores and choosing which information would be transferred to the next type of store. In the preceding example, fourth-grade students were able to grasp a rather complex model of cognition by breaking it down into simple, concrete parts and physically acting out the interaction between these parts. In another example, fifth graders grasped the idea of genotypes and phenotypes by imagining the marriage of a brown-haired princess (with a dominant hair color gene) and blue-haired monster (with a recessive hair color gene), and drawing simple Punnett squares to map out what color hair their children would have.

Undergraduate Student Feedback on Outcomes

At the conclusion of the course, we encouraged, but did not require, the undergraduate students to complete

---

1 Additional procedural details and results can be found at http://www.stolaf.edu/depts/psych/faculty/gary/sts/index.html.
Table 1. Undergraduate Student Presentation Evaluation Results

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Agree Somewhat</th>
<th>Disagree Somewhat</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don't Know/Doesn't Apply</th>
<th>All &quot;Agree&quot; Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was a valuable learning experience</td>
<td>42.33</td>
<td>39.22</td>
<td>13.20</td>
<td>2.14</td>
<td>1.36</td>
<td>1.55</td>
<td>0.19</td>
<td>94.76</td>
</tr>
<tr>
<td>Was an enjoyable experience</td>
<td>59.57</td>
<td>29.78</td>
<td>9.07</td>
<td>0.99</td>
<td>0.00</td>
<td>0.20</td>
<td>0.26</td>
<td>98.42</td>
</tr>
<tr>
<td>Helped increase my understanding of psychology</td>
<td>26.74</td>
<td>35.08</td>
<td>27.33</td>
<td>5.62</td>
<td>2.52</td>
<td>2.13</td>
<td>0.58</td>
<td>89.15</td>
</tr>
<tr>
<td>Should be a regular part of this course</td>
<td>48.26</td>
<td>32.36</td>
<td>10.27</td>
<td>2.71</td>
<td>2.13</td>
<td>2.52</td>
<td>1.74</td>
<td>90.89</td>
</tr>
</tbody>
</table>

an online course evaluation form on which they rated four questions regarding the school presentations on a 7-point scale (see Table 1). Beginning in Fall 2006, students also responded to the following open-ended request: “Please comment on HOW you think the Elementary School Presentation project improved your understanding of psychology (if applicable).”

**Elementary School Feedback on Outcomes**

We provided a packet for each group to give to their elementary teacher containing a teacher feedback form and survey forms for the elementary students. The teacher’s feedback form contained seven questions with responses on a 5-point scale (see Table 2), as well as a space for any comments. Elementary students (kindergarteners to fifth graders) circled the face (sad, happy, or neutral) on their feedback forms that best reflected how they felt about the presentation.

**Grading**

The presentation component of the course accounted for 25% of undergraduate students’ final grade, and the presentation grade consisted of three components: (a) feedback from teachers and students; (b) peer assessment of participation in planning and conducting the presentation; and (c) a short, three-page reflective paper (written individually) describing what they did in their presentation and justifying their specific presentation content by incorporating primary source literature.

**Evaluation Results**

**Undergraduate Student Evaluations**

Data from the online course evaluation forms for a total of 518 undergraduate college students (167 men and 351 women) over seven semesters showed extremely favorable responses to the project (Table 1). Importantly, the earlier the student was in his or her college experience when he or she completed the presentation, the greater the positive impact the student reported it having on his or her understanding of psychology, $r_s(505) = -.15$, $p < .001$. In addition, qualitative responses consistently demonstrated that students were aware that, as one student stated, “Challenging yourself to teach a fairly complex scientific concept to elementary students forces you first, to understand the concept inside and out, and second, to be creative in designing ways to teach and explain the material, which further cements the concept for you.”

**Elementary Teacher and Student Evaluations**

Feedback from the elementary school teachers ($N = 199$) over seven semesters on 5-point ratings scales showed very positive responses across all items (Table 2). Of particular importance, teachers were very interested in participating in the project again in the future, and many have been involved on multiple occasions. In addition to the ratings, teachers consistently offered very useful advice for improvement. Elementary student feedback showed that students also greatly enjoyed the presentations (Table 2).
Conclusions and Future Directions

Quantitative and qualitative data regarding evaluation of this EL “students teaching students” project suggested that the project’s goals are being met with regard to the satisfaction reported by both the undergraduate students and the elementary schools. Undergraduate reflections on the project through papers and questionnaires showed an awareness that communicating psychological concepts to elementary school students calls for them to “translate” these concepts into age-appropriate language for their audience and that this requires the presenter to have a deeper and more flexible understanding of the concept itself. It is worth noting that freshmen students reported getting the most academic benefit from the presentation, suggesting that these EL opportunities should be offered as early in the psychology curriculum as possible.

As this is an ongoing project, our analysis of evaluation results and assessment of learning outcomes are constantly under revision. Most recently, we have begun to examine quantitative ways to better evaluate both undergraduate and elementary student learning. One possible evaluation method for undergraduate student learning would be comparing students’ final exam performance on questions concerning the material they presented against the performance of classmates who did not present that material. To assess elementary school learning, an age-appropriate quiz could be developed by each presentation group and administered to the elementary students as a pre- and posttest on the day of the presentation.

This project represents an EL opportunity for large undergraduate classes with a number of benefits for both undergraduate and elementary students that are realized in a mutually stimulating and enjoyable way. As one undergraduate student stated, “In order to really know a subject, one should try to teach it. The elementary school project was a blast!”

References


Table 2. Teacher and Elementary Student Presentation Evaluation Results

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Was the group prepared? (1 = unprepared; 5 = well prepared)</td>
<td>4.70</td>
<td>0.14</td>
</tr>
<tr>
<td>Q2. Based on your observations, how well do you think your students enjoyed the presentation? (1 = not at all; 5 = a great deal)</td>
<td>4.66</td>
<td>0.10</td>
</tr>
<tr>
<td>Q3. Did the student presenters display a clear understanding of the presentation material? (1 = not at all; 5 = a great deal)</td>
<td>4.66</td>
<td>0.09</td>
</tr>
<tr>
<td>Q4. How interactive was the presentation? (1 = not interactive; 5 = very interactive)</td>
<td>4.48</td>
<td>0.13</td>
</tr>
<tr>
<td>Q5. Was the material presented at an age-appropriate level? (1 = inappropriate; 5 = very appropriate)</td>
<td>4.53</td>
<td>0.11</td>
</tr>
<tr>
<td>Q6. Based on your observations, how much did your students learn from the presentation? (1 = very little; 5 = a great deal)</td>
<td>4.06</td>
<td>0.13</td>
</tr>
<tr>
<td>Q7. How interested would you be in working with the St. Olaf Psychology Department on this project in the future? (1 = not interested; 5 = very interested)</td>
<td>4.69</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Elementary student responses

| Did you like the presentation? (proportion of smiling faces to neutral and sad faces/5) | 4.20| 0.11 |

Note. N = 199. Q2 was introduced in Spring 2007 (n = 83).


Notes

1. Gretchen J. van der Linden is now at the School of Social Work, University of Minnesota, Twin Cities.

2. This project could not have happened without the support of Northfield elementary school principals Scott Sannes (Sibley Elementary), Nancy Antoine (Bridgewater Elementary), and David Craft (Greenvale Park Elementary); former principals Diane Kinneberg, Julie Nielsen, and Jeff Roland; and the many teachers who participated. Dan Forstner, Department of Education, St. Olaf College, has provided advice and assistance for the project in addition to invaluable guest lecturing on presenting to elementary school-age children. Natalie Wall, Reference Librarian, St. Olaf College, has presented an information literacy class to help students find sources in the library to prepare them for finding the references required for students’ presentation paper. The class TAs who assisted in implementing the project were Katie Johnson, Elizabeth “Biz” Mills, Kristyn Aasen, Brian McCormick, Rachel Varland, Laura Dingley, Hannah Thiesen, Laura Barnard, Gretchen van der Linden, Kirsten Meyer, Abby Hughes, Vanessa Brown, and Shannon Thornblad.

3. Portions of this article were presented at the Innovations in the Scholarship of Teaching and Learning at the Liberal Arts Colleges conference (February 2007) in Northfield, MN.

4. Send correspondence to Gary M. Muir, Department of Psychology, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057; e-mail: muir@stolaf.edu
We analyzed multiple-choice questions provided in test banks for introductory psychology textbooks. In Study 1, we found that about 70% of students responded correctly to a given item, and there was a significant but inconsistent difference in difficulty across chapter topics. In Study 2, a single class of introductory psychology students took comprehensive exams of randomly selected test items that varied in estimated difficulty. As before, chapter topic had a significant, yet inconsistent effect, but test items had several desirable properties compared to Study 1. These data allow several recommendations for instructors who want to improve the characteristics of the multiple-choice exams they use.

In a typical introductory psychology course, students attempt to learn about a large number of concepts—perhaps as many as several thousand (Zechmeister & Zechmeister, 2000). It is unlikely that these concepts are equally difficult to master, and an accurate assessment of content difficulty is important to curriculum development, implementation, and evaluation. In the studies reported here, we investigated students’ exam performance as an indicator of content difficulty. Specifically, the goals of this study were to determine the psychometric difficulty and discriminability of test items used in introductory psychology courses and to compare these indexes of content difficulty across topic areas in the course as typically offered.

Although there are various approaches to the operationalization of test item difficulty, one of the more accessible definitions is the proportion of students giving the correct answer to an item (e.g., Rogers, 1995). An item passed by too few or too many students may not contribute useful information about what students have learned, and test items should have a range of psychometric difficulty centered on .50. Difficulty estimates are readily obtainable today because testing of many introductory psychology students uses multiple-choice items (Irion, 1976) that can be scored with optional item analyses. The widespread use of multiple-choice testing also increases the ecological and face validity of estimates of content difficulty based on such test results.

Item analysis of difficulty can be paired with an assessment of discriminability, a test item’s power to separate students who score well from those who score poorly on a test. There are several operational definitions of discriminability, including the point-biserial correlation of items and total test scores and the difference between the proportions of high scorers and low scorers who answer an item correctly. Difficulty and discriminability estimates can be combined to facilitate revision and refinement of objective examinations and to guide decision making about lecture content, reading assignments, and other aspects of the curriculum. For example, an instructor might choose to omit test items unless at least 25% of the class pass them and they have a discriminability of at least .10. As well, under the assumption that limited resources demand selectivity in curriculum development, those areas of greater difficulty would be stronger candidates for resource allocation.

Commercial test-item banks treat item analysis inconsistently. In most cases, they do not detail the construction or analysis of test items and, when difficulty estimates are given, there is often a lack of information concerning how these ratings were derived. Among the small set of texts we have examined, Kalat and Stonebraker (1996) gave the most complete data for the test bank accompanying Kalat (1996). Many of their test items have been evaluated for difficulty and discriminability at a single, public, postsecondary institution in the United States.

In the studies reported here, we gathered data on several thousand test items selected from test banks used by a convenience sample of instructors teaching introductory psychology at both colleges and universities. We analyzed the items’ difficulty and discriminability by topic area following the taxonomy of Griggs, Jackson, and Napolitano (1994). In doing so, we wanted to determine if exam performance varied systematically by topic area and, if so, to consider the implications of these differences for the teaching professor.

**Study 1**

**Method**

Items used for the analysis came from publishers’ test banks that were provided for six introductory psychology
textbooks (Atkinson, Atkinson, Smith, Bem, & Nolen-Hoeksema, 1996; Baron, Earhard, & Ozier, 1998; Carlson & Buskist, 1997; Kalat, 1996; Morris, 1996; Sternberg, 1995, 1998). Griggs (1999) recently rated these texts in the low-middle to high range of difficulty. The texts were used in various institutions offering the course, taught in both one-semester and two-semester formats by nine different instructors. The instructors were a convenience sample comprised of about equal numbers of men and women between the ages of 30 and 50 years who had specialization in a range of subdisciplines including social psychology, perception, cognition, psychophysics, learning, and neuroscience. These instructors were part of a province-wide effort to characterize teaching challenges in introductory psychology and develop technological enhancements to the curriculum, and they volunteered following a provincial annual meeting on the teaching of psychology. The data represent responses of 2,631 students to more than 4,000 questions. A summary of the sample appears in Table 1.

Participating instructors submitted their test results in a variety of formats. We coded the data following the 18-topic chapter taxonomy used previously by Griggs et al. (1994). The analyses that follow do not include items explicitly taken from chapters on Human Sexuality and Applied Psychology, bearing in mind that some applied questions appeared in other chapters. We coded items that did not have both difficulty and discrimination scores as missing data.

Results

Two indexes form the basis of the analyses to follow. Difficulty indicates the proportion of students who responded correctly to an item (see Rogers, 1995). Discriminability is an index of an item's ability to separate those who performed well on a test from those who did not. We calculated discriminability as \( p_{\text{High}} \) minus \( p_{\text{Low}} \), where \( p_{\text{High}} \) is the proportion of students in the highest quartile passing the item and \( p_{\text{Low}} \) is the proportion of students in the lowest quartile passing the item.

Average difficulty and discriminability data are shown by chapter topic in Table 2, in which the analysis eliminated items that had a difficulty of less than .25 or a discriminability less than .10. We used this screening in response to suggestions that difficult items with such poor discrimination are not useful in the context of introductory psychology teaching and are associated with lower test reliability (Hopkins, 1998). Table 2 shows that, on average, about two thirds to three fourths of the students passed an item. Discriminability averaged .36, a level that is limited because average difficulty is lower than .5, the point at which discriminability is maximal (Hopkins, 1998, p. 257). Collapsed across topic, 90% or more of students passed almost 9% of the test items. Almost 15% of the items had discriminability values of .20 or less.

We also examined the difficulty data in relation to several variables available in the data set. As might be expected on the basis of sample size alone, these effects were significant but probably more important, they accounted for little variance in the data. There was a main effect of chapter topic, \( F(15, 3322) = 6.25, p < .001 \), but it accounted for only 3.1% of the variance. There was a main effect of textbook, \( F(6, 3331) = 17.77, p < .001 \), which accounted for only 3.1% of the variance in the difficulty data. Items from the test bank used with Baron et al. (1998) were least difficult (\( M = .77 \)), and items associated with Sternberg (1998) were most difficult (\( M = .65 \)). We evaluated item difficulty in relation to textbook difficulty as estimated by Griggs (1999) and, although the main effect was significant, \( F(2, 3335) = 37.01, p < .001 \), only 2.2% of the variance was accounted for by the effect. Furthermore, texts of intermediate difficulty yielded

### Table 1. Descriptive Data on Test Item Sample

<table>
<thead>
<tr>
<th>Textbook</th>
<th>Textbook Level ( a )</th>
<th>Type of Institution</th>
<th>Length (1 or 2 Semesters)</th>
<th>No. of Items</th>
<th>Questions From Test Bank</th>
<th>No. of Instructors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baron, Earhard, &amp; Ozier (1998)</td>
<td>Middle</td>
<td>University</td>
<td>1</td>
<td>248</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Carlson &amp; Buskist (1997)</td>
<td>High</td>
<td>University</td>
<td>1</td>
<td>148</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Kalat (1996)</td>
<td>Middle</td>
<td>University</td>
<td>1</td>
<td>563</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>Morris (1996)</td>
<td>Low-Middle</td>
<td>College</td>
<td>2</td>
<td>1,770</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Sternberg (1995)</td>
<td>High</td>
<td>College</td>
<td>2</td>
<td>330</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Sternberg (1998)</td>
<td>High</td>
<td>University</td>
<td>1</td>
<td>513</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>1</td>
<td>4,025</td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

\( a \)As evaluated by Griggs (1999).

### Table 2. Difficulty and Discriminability by Topic: Study 1

<table>
<thead>
<tr>
<th>Topic</th>
<th>Difficulty</th>
<th>Discriminability</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>.703</td>
<td>.338</td>
<td>172</td>
</tr>
<tr>
<td>Methods</td>
<td>.713</td>
<td>.343</td>
<td>80</td>
</tr>
<tr>
<td>Biological</td>
<td>.688</td>
<td>.365</td>
<td>301</td>
</tr>
<tr>
<td>Sensation and perception</td>
<td>.684</td>
<td>.353</td>
<td>340</td>
</tr>
<tr>
<td>Consciousness</td>
<td>.719</td>
<td>.337</td>
<td>148</td>
</tr>
<tr>
<td>Learning</td>
<td>.671</td>
<td>.375</td>
<td>229</td>
</tr>
<tr>
<td>Memory</td>
<td>.719</td>
<td>.353</td>
<td>194</td>
</tr>
<tr>
<td>Cognition</td>
<td>.635</td>
<td>.357</td>
<td>141</td>
</tr>
<tr>
<td>Intelligence</td>
<td>.696</td>
<td>.375</td>
<td>80</td>
</tr>
<tr>
<td>Motivation</td>
<td>.708</td>
<td>.378</td>
<td>223</td>
</tr>
<tr>
<td>Development</td>
<td>.742</td>
<td>.360</td>
<td>439</td>
</tr>
<tr>
<td>Personality</td>
<td>.721</td>
<td>.362</td>
<td>182</td>
</tr>
<tr>
<td>Stress</td>
<td>.722</td>
<td>.352</td>
<td>105</td>
</tr>
<tr>
<td>Abnormal</td>
<td>.697</td>
<td>.374</td>
<td>314</td>
</tr>
<tr>
<td>Therapies</td>
<td>.665</td>
<td>.428</td>
<td>218</td>
</tr>
<tr>
<td>Social</td>
<td>.689</td>
<td>.370</td>
<td>172</td>
</tr>
<tr>
<td>Overall M/Total</td>
<td>.698</td>
<td>.365</td>
<td>3,338</td>
</tr>
</tbody>
</table>
easier questions (M = .74) compared to either easier (M = .67) or more difficult texts (M = .69). Finally, we asked whether test items given to college students (M = .68) differed in difficulty compared with test items given to university students (M = .72). The effect of institution was significant, F(1, 3336) = 60.72, p < .001, but only 1.8% of the variance was related to the factor.

Near the completion of these analyses, we received additional data from another instructor in the research group who had used Gray (1994) and Weiten (1998) in teaching the course. Because the measure of discriminability, the point-biserial correlation between each item and total test scores, is not commensurate with the discriminability measure used earlier, we did not think it appropriate to combine the data. Instead, we present them as a substantive replication of the previous study.

Test items were from the test banks provided with the text. After applying the same criteria used previously, that is, selecting only those items that had a difficulty greater than .25 and a discriminability greater than .10, there were 1,315 items in the data set.

Although the data came from a different instructor using different texts at different institutions, the trends were similar to those already reported. Across all chapter topics, average difficulty was approximately .70 and average discriminability near .30. The range of difficulty (approximately 10%) and discriminability (approximately 13%) indicated similar dispersion in the two data sets. The main effect of chapter topic was significant, F(15, 1299) = 1.99, p = .013, but, as with the initial data, it accounted for a small (2.3%) amount of the variance in the data. In fact, the correlation between average difficulty in the two data sets was zero.

**Study 2**

Study 2 offered a consistent picture of the objective difficulty of multiple-choice tests given to introductory psychology students. Performance on test items did not differ systematically as a function of the materials from which they derived, at least at the rather coarse, chapter level. One hypothesis that might account for this consistency is that instructors may intentionally select items that are similar in difficulty. If this type of selection occurs and if the test material does indeed vary in difficulty as a function of topic area, then a random selection of test items would produce a pattern reflecting this relative difficulty. This thinking formed the basis for Study 2. Another purpose of the study was to determine whether psychometric difficulty was related to the 3-point difficulty ratings provided in the test bank.

**Method**

The items for Study 2 came from the test bank associated with Sternberg’s (1998) text. A single class responded to one of the multiple-choice examinations referred to in the previous study. For comparison, the difficulty and discriminability of items were examined for the sample of students who completed both versions of the final examination. Item difficulty and discriminability were not significantly different across the two tests.

**Results**

Average difficulty and discriminability appear by chapter topic in Table 3, in which the sample sizes reflect the exclusion of items that did not meet the criteria of having a difficulty value greater than .25 and a discriminability greater than .10. As in Study 1, these selection criteria resulted in a considerable reduction in the data set: A total of 22% of the items were eliminated.

The most salient trends in the data are that average difficulty, at .60, was approximately 10% lower (that is, the items were more difficult) in comparison to results obtained with instructor-selected items. In addition, the average difficulty was approximately 5% lower than from the same instructor, using the same text, when he selected test items nonrandomly. Understandably, discriminability was increased somewhat.

A one-way ANOVA on item difficulty produced a significant effect of chapter topic, F(15, 218) = 1.82, p = .033, that accounted for 11% of the variance in the data. We also analyzed the difficulty scores as a function of question type and estimated difficulty. Factual questions (M = .61) were not easier than conceptual questions (M = .59), p = .14, effect size = .01. Easy questions (M = .63) were answered correctly more often than questions of nominally intermediate difficulty (M = .60) and these, in turn, were answered correctly more often than difficult items (M = .56). The effect of nominal difficulty was, therefore, in the anticipated direction, but was only marginally significant (p = .073) and accounted for only 2% of the variance in the data. The interaction of question type and difficulty was not significant (p = .50).

Parallel analyses on the discriminability data indicated that the effect of chapter topic was nonsignificant (p = .59). The analysis of question type and nominal difficulty revealed

<table>
<thead>
<tr>
<th>Topic</th>
<th>Difficulty</th>
<th>Discriminability</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>.568</td>
<td>.378</td>
<td>12</td>
</tr>
<tr>
<td>Methods</td>
<td>.690</td>
<td>.367</td>
<td>13</td>
</tr>
<tr>
<td>Biological</td>
<td>.563</td>
<td>.425</td>
<td>12</td>
</tr>
<tr>
<td>Sensation and perception</td>
<td>.517</td>
<td>.392</td>
<td>20</td>
</tr>
<tr>
<td>Consciousness</td>
<td>.644</td>
<td>.378</td>
<td>14</td>
</tr>
<tr>
<td>Learning</td>
<td>.595</td>
<td>.295</td>
<td>13</td>
</tr>
<tr>
<td>Memory</td>
<td>.555</td>
<td>.388</td>
<td>12</td>
</tr>
<tr>
<td>Cognition</td>
<td>.561</td>
<td>.371</td>
<td>19</td>
</tr>
<tr>
<td>Intelligence</td>
<td>.657</td>
<td>.440</td>
<td>13</td>
</tr>
<tr>
<td>Motivation</td>
<td>.663</td>
<td>.415</td>
<td>11</td>
</tr>
<tr>
<td>Development</td>
<td>.649</td>
<td>.364</td>
<td>25</td>
</tr>
<tr>
<td>Personality</td>
<td>.570</td>
<td>.480</td>
<td>11</td>
</tr>
<tr>
<td>Stress</td>
<td>.671</td>
<td>.398</td>
<td>12</td>
</tr>
<tr>
<td>Abnormal</td>
<td>.640</td>
<td>.436</td>
<td>6</td>
</tr>
<tr>
<td>Therapies</td>
<td>.509</td>
<td>.439</td>
<td>14</td>
</tr>
<tr>
<td>Social</td>
<td>.559</td>
<td>.394</td>
<td>25</td>
</tr>
<tr>
<td>Overall M/Total</td>
<td>.596</td>
<td>.393</td>
<td>234</td>
</tr>
</tbody>
</table>
that there was no difference ($p = .79$) in discriminability between factual ($M = .40$) and conceptual questions ($M = .39$). There was also not an interaction of question type and nominal difficulty ($p = .68$). There was, however, a main effect of nominal difficulty, $F(2, 228) = 3.39, p = .04$, that accounted for $3\%$ of the variance. Easy items ($M = .40$) and items of intermediate difficulty ($M = .41$) had greater discriminability than difficult items ($M = .35$).

To examine the consistency of the main effect of chapter topic, we correlated the data from Study 2 with the two data sets in Study 1. In the first case, $R^2 = .30$. In the latter case, $R^2 = .02$. Even the squared correlation between the instructor-selected items and random-selected items from the same instructor using the same text was only .24.

Discussion

The results of Study 2, like those of Study 1, indicate that on the whole, test items taken from commercial test banks have poor psychometric properties. There is evidence that one in five items were inappropriate for testing purposes. However, the data suggest that, after removal of questionable items, Study 2 produced a test in which difficulty was normally distributed about a value that is close to the optimal .50 (Hopkins, 1998, p. 257) and was neutral with respect to content area. Discriminability was relatively good and did not differ across topic areas. Discriminability may have been lower than desired because average difficulty departs from the optimal value and because difficult test items tended to be answered incorrectly by even the higher scoring students.

Evaluation of the effects of nominal difficulty and question type produced several provocative results. First, although there was a significant effect of nominal difficulty, it accounted for a small proportion of the variance and reflected a scant $7\%$ difference in mean psychometric difficulty across the three levels of nominal difficulty. It would be desirable psychometrically to produce test items that have a greater spread of difficulty. Second, factual and conceptual questions produced the same difficulty and discriminability. One might expect that conceptual questions would be more difficult to answer and would serve to discriminate the higher scoring and lower scoring students. Thus, in at least two respects, the nominal classifications of test items may be of questionable validity.

General Discussion

This analysis yields a somewhat mixed picture concerning the nature of the items used to assess learning in introductory psychology classes. There may be reason for concern that items are less difficult than is psychometrically desirable and that, partly as a consequence, their discriminative power is reduced. The range in difficulty by chapter topic (about $12\%$ in Study 1 and $19\%$ in Study 2) does not seem large and is not consistent, but when expressed in terms of a letter grade it is substantial and might be a concern for individual instructors who are assigning grades.

On a more positive note, there was an almost uncanny consistency to the test items presented to introductory psychology students when the items were instructor selected. Regardless of topic area, about $70\%$ of students were able to answer test questions correctly. This is a norm that held across chapter topics, textbook authors, and estimated textbook difficulty (Griggs, 1999), in that none of these factors accounted for more than about $3\%$ of the variance in difficulty scores.

These conclusions should be tempered by the knowledge that, in common with most naturalistic studies of teaching, many extraneous variables affect test performance. Among the most important influences are those related to teachers and teaching method, the individual test items, and student characteristics. On the other hand, these data come from a process that emphasized ecological validity and mirrored common practice. In addition, the data are consistent with our personal experience as instructors. In creating tests, we tend to pick a few items that everyone will pass (the confidence builders), a few items that only a few will answer correctly (to separate the A and B students), and the rest so that they are neither obvious nor impossible.

Study 2 produced test items that were preferable in some ways to those of Study 1. Perhaps this difference reflects the comprehensive nature of the exams given in Study 2. Comprehensive exams that cover semester-long material are more difficult than noncomprehensive exams and may, therefore, have better psychometric properties. On the other hand, the differences observed might be due to across-study differences in the method of item selection. Both approaches yielded about the same proportion of test items that students answered at levels worse than chance or that failed to discriminate high and low test scorers. Both approaches produced significant but inconsistent effects of chapter topic that may be important at the instructor level but do not have broader curricular implications. However, Study 2 items had better distributional properties and tested the content more thoroughly than when item selection allowed for instructor biases. Additionally, students may be motivated to read the text more thoroughly if they know that a larger population of test items is randomly selected.

On the other hand, random selection of items may assess relatively unimportant facts or applications. Instructors who select items themselves have the option of weighting the test toward what they believe are important concepts or toward material covered explicitly in lecture, tutorials, and labs. Finally, there may be good pedagogical reasons to increase the number of students answering test items correctly, and items selected by instructors appear to achieve this goal reliably.

We tentatively offer several recommendations to those teaching and developing test materials for the first course in psychology or, in fact, for any course in which multiple-choice items evaluate student learning. Carry out item analyses on your own tests and set selection criteria before administering the test. For example, one might construct the test to have $5\%$ of items passed by $90\%$ of students to boost confidence, $5\%$ of items passed by $10\%$ of students, and the remainder selected to have an average difficulty of .50. Send the results of your item analyses to test bank developers, along with data about yourself as instructor and the students who provided the data. Ask test bank writers to provide detailed item analyses. Relatedly, develop or ask test bank developers to create test items classified by difficulty and that have a broader range of mean diffi-
culty than is presently the case. Consider using random selection of test items in cumulative examinations, particularly if there is a desire to increase test difficulty, motivate students to read the text, and cover the material more exhaustively. Of course, completely random selection allows the possibility that several items would ask essentially the same question, and so some limitations to completely random item selection are desirable. Finally, regularly inform students of the test construction process, so that they understand the genesis of their grades, recognize your concern for developing a good test of their learning, and see how psychologists use their expertise in real-world settings.

References


Notes

1. This research was funded by a grant from the Province of Alberta’s Learning Enhancement Envelope.
2. We thank Scott Oddie, Genevieve Thurlow, Cindy Lahar, Suzanne Hala, John Mueller, Elbia Slawinski, Susan Graham, Lenora Brown, and Karsten Loepelman, who provided data for Study 1.
3. Send correspondence to C. T. (Chip) Scialfa, Department of Psychology, University of Calgary, Calgary, AB, Canada T2N 1N4; e-mail: scialfa@ucalgary.ca.
In this article I explore the educational value of students writing questions on what they do not know about course content. Students bring questions to class on assigned chapters and we discuss the questions. Later, students write one question for each exam. Students in 10 classes agreed that the technique helped them see alternative ways of viewing the course material, understand the limitations of texts and lectures, and relate current information to previously learned material. This technique is applicable to both introductory and advanced classes.

The concept of ignorance plays an important function in scientific disciplines. In part, this is because young sciences, such as psychology, have not yet developed an indisputable set of methods and facts. More fundamentally, all scholarly endeavor begins with ignorance and curiosity about the unknown (Stocking, 1992; Witte, Kerwin, & Witte, 1988). If teachers of psychology are to impart to students not just the facts of the discipline, but also the process by which knowledge is created, we must increase their awareness of the role of ignorance in psychology.

One way to accomplish this goal is to encourage students to recognize the limitations of textbooks. Textbook authors face some difficult decisions in deciding what to include in their texts. Consequently, authors sometimes leave out information that may be reasonably construed as relevant to the discussion. For instance, LeUnes (1983) reviewed abnormal psychology texts and found that although most discussed Watson and Rayner’s (1920) research on Little Albert, only 4 of 27 texts discussed what became of Albert after the study.

Instructors may teach students to recognize omissions in textbooks by asking them to write ignorance questions. An ignorance question is a question stimulated by class material but not directly answered by it (Peden & Keniston, 1991; Witte et al., 1988). For example, one of my introductory students asked “Is there a limit to the number of memories we can repress?” Although the text discussed capacity limitations in short-term memory and the idea of repression, it did not integrate the two concepts. Thus, the question was relevant to but went beyond the text material. Some of the queries that make good ignorance questions are those that draw implications from the subject matter, identify applications of the subject matter, or draw relations between current concepts and previously learned material.

Writing ignorance questions shares some features of techniques that promote active learning, critical thinking, and integration of material. Instead of memorizing the material verbatim, students must think about the content. For example, active processing of material at a deeper semantic level facilitates retention (Mathie et al., 1993). The activity also requires students to think about course material in a different way, which improves critical thinking (Budesheim & Lundquist, 1999; Gorman, Law, & Lindegren, 1981). In addition, formulating an ignorance question encourages students to relate new concepts to those learned previously, thereby creating links among related concepts.

In this study I examined the value of using ignorance questions in a variety of classes. Peden and Keniston (1991) found that students in an introductory psychology class wrote increasingly better ignorance questions throughout the term, and students endorsed the use of the technique for future classes. Similarly, Gray (1993) found that most introductory students liked to write questions after they got the knack for doing so. Using a very different group of students, Witte et al. (1988) reported that third- and fourth-year medical students favorably viewed a curriculum based on medical ignorance.

I have used ignorance questions in 10 classes over several years. These include classes in introductory psychology (honors and regular sections) and senior-level classes in psychology of language (three times), language development (twice), learning and behavior modification (twice), and child development. The procedure used was the same in different classes with slight variations.

Method

I introduce students to the idea of ignorance questions on the first day of class and I provide a handout. I define the term, emphasizing the distinction between an ignorance question and other types of questions (e.g., questions that ask for clarification of the material presented in a textbook). In addition, I discuss how we will use these questions in our class. I inform students that I will ask them, at intermittent intervals, to bring a question to class and to write one ignorance question for each exam.

In general, I ask students to bring an ignorance question two or three times per semester. When I first used this technique, I had students read their questions to the class and then we discussed them. More recently, I have had students write questions on pieces of paper without their names and hand them to me at the outset of class. My observation is that in this less threatening environment, students write better questions.

When we first discuss their questions, I typically read some or all of the questions aloud and explain which ones are good.
ignorance questions and which are not. This activity often leads to a discussion of the criteria I use for grading ignorance questions (discussed subsequently). We also discuss strategies for generating questions. For instance, I encourage students to think of questions as they read, rather than waiting until the end of a chapter to write a question.

Over time, discussion moves more to the content of the questions, including divergent perspectives, implications, or issues not adequately explored in the text. Often, I read all the questions and then ask students which ones they want to discuss. Initially I do most of the answering. By spontaneously trying to answer their questions, I am modeling an activity I want to encourage in them: the ability to bring one’s existing knowledge to bear on a new problem. Thus, I encourage students to view instructors, and themselves, as active problem solvers rather than as “walking encyclopedias.” Ordinarily, the discussion of ignorance questions takes about 15 to 20 min of class time.

Students also write one ignorance question for each exam. I encourage them to bring prewritten questions to the exam. I grade these questions on a 5-point scale (modified from Peden & Keniston, 1991). I give a 5 to questions that are relevant to the subject matter but are not answered in the text. Many excellent ignorance questions take the form of if–then statements (e.g., “If all humans express emotions with the same facial expressions, then why is there so much distrust and mis-understanding between cultures?”). Good questions that are not entirely clear (e.g., “Do dreams reflect a person’s personal-ity or are they reflective of experience?”) are scored as a 4. I score questions answered in the text or that presuppose a mis-understanding of text material (e.g., “If it is true that the two sides of the brain don’t communicate with each other …”) as a 3. Questions that are not relevant to the assigned content are graded as a 2. Finally, I grade jokes or nonserious questions (e.g., “Why are all men such dogs?”) as a 1.

Results

I assessed the effectiveness of the technique in two ways: student evaluation of the technique and student scores on questions written for exams. I identified student reactions to this activity by using a 10-item evaluation form at the end of the semester. Some items were used with all of the classes (N = 134) and others only in some. Table 1 presents the responses for introductory and senior-level students. Both groups of students agreed that the technique helped them to write good questions about the course material, to better understand the limitations of texts and lectures, to see alternative ways of interpreting or applying the course material, and to relate current material with previously learned material. There was less agreement that the task of writing questions was difficult and disagreement with the statement that it is more important to be able to answer questions than to ask questions. Overall, students recommended that this activity be used in the future in the class.

The results in Table 1 are similar for introductory honors students and for senior-level students. Introductory students (both honors and regular sections) were less likely than advanced students to agree that it was difficult to write questions. I have found that seniors complain more about the difficulty of writing a good question, at least initially. Writing questions may be more difficult for advanced students because it requires skills not often utilized in their otherwise successful academic careers.

The results vary from those of Peden and Keniston (1991), who reported that students found writing ignorance questions to be initially difficult, but improved sub-

<table>
<thead>
<tr>
<th>Item</th>
<th>Introductory*</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helped me write good questions</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>Writing questions was difficult</td>
<td>3.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Better understood limitations of text</td>
<td>3.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Clarified misunderstandings about text</td>
<td>3.4</td>
<td>1.0</td>
</tr>
<tr>
<td>Encouraged alternative ways of interpreting material</td>
<td>4.0</td>
<td>0.8</td>
</tr>
<tr>
<td>Helped relate information in different chapters</td>
<td>3.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Amount of class time was appropriate</td>
<td>4.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Inclusion on exams was appropriate</td>
<td>4.1</td>
<td>0.8</td>
</tr>
<tr>
<td>More important to answer questions than to ask them</td>
<td>2.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Recommend activity be used in the future</td>
<td>4.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Note. Students rated statements based on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree).

*Students rated statements based on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Discussion

The results suggest that students in a variety of classes have found writing ignorance questions to be a valuable experience. It is easy for students to learn to write good questions. It is relatively easy to learn how to score them as well; if desired, instructors could train teaching assistants to grade ignorance questions.

The results vary from those of Peden and Keniston (1991), who reported that their students found writing ignorance questions to be initially difficult, but improved sub-

Vol. 28, No. 2, 2001
stantially through the term. In contrast, my students did as well at the beginning and the end of the term. One reason for the different results may be that my students had experience in writing questions in class before the first exam and the class discussion emphasized the criteria for good questions. Many students seemed to grasp the idea of ignorance questions right away. In addition, many of my students were in advanced classes, whereas Peden and Keniston’s students were introductory students.

As noted earlier, asking questions promotes active learning, a concept endorsed by many instructors (e.g., Benjamin, 1991). Studies indicate that writing multiple-choice and essay questions promotes test performance in introductory students (Foos, 1989; Kerkm, Kellison, Piñon, Schmidt, & Lewis, 1994). In contrast, although ignorance questions may be effective in teaching content, they may be most useful in promoting intellectual curiosity.

Perhaps the main benefit is the qualitative change in classroom ambience. My impression has been that students ask more questions during class and that their questions and comments are progressively more reflective throughout the term. Although students are initially required to write questions for class discussion, this requirement can be eliminated midway through most terms and students continue to be active in class discussions. On the cost side, the amount of time devoted to ignorance questions detracts from lecture time, although the loss of time is minimal. Additionally, because the definition of a good ignorance question is something that the text does not mention, this technique forces experienced instructors to read their textbooks carefully! When students write questions that refer to material that is commonly in textbooks, instructors need to determine whether it is in the current textbook.

In short, this is an easy-to-use technique that opens up class discussions in an appealing way, and students rate the technique favorably. It can be a valuable component of any course that encourages students to be more thoughtful and critical.

References


Notes

1. An earlier version of this article was presented at the American Psychological Society Institute on the Teaching of Psychology, Washington, DC, June 1994.

2. I thank Blaine F. Peden, Ruth Ault, Todd Jackson, and three anonymous reviewers for their thoughtful comments on an earlier draft of this article and Kristin Ericson Foster for assisting with the interrater reliability analysis.

3. Send correspondence to David W. Carroll, Psychology Program, University of Wisconsin, Superior, WI 54880; email: dcarroll@staff.uwsuper.edu.
unanimous in reporting that the course met or exceeded their expectations. Future special topic courses focusing on different careers can add to the information in our career library while giving more students the opportunity to explore careers with a master's degree.

**References**


**Notes**

1. We thank all the students enrolled in Psychology 398: Careers in Psychology during the fall semester, 1998–99. Because of their hard work, many of our undergraduates gained a greater appreciation of career opportunities in psychology.

2. Send correspondence to Eugene B. Zechmeister, Department of Psychology, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, IL 60626; e-mail: ezechme@luc.edu.

---

**The Use of Discrimination Indexes in Constructing Course Exams: A Question of Assumptions**

**Daniel R. Stalder**  
**University of Wisconsin–Waukesha**

Some instructors use the discrimination index (or item-total correlation) to assess their exams’ reliability. These instructors may exclude items with near zero or negative indexes from an exam (or later administrations of the exam) to improve reliability. This article examines the potential pitfalls of this practice and questions the underlying statistical assumption that there is a single construct that course exams should measure.

Similar to researchers who strive to develop measures that are both reliable and valid, many instructors try to write reliable and valid course exams. At a minimum, most instructors probably try to write items that adequately reflect material from lectures and readings, representing content validity. Some instructors also use the discrimination index (computed per item) to assess reliability. This statistic is conceptu-
ally similar to the item-total correlation and, if positive for an item, indicates that the top portion (e.g., one third) of exam scorers were more likely to answer that item correctly, relative to the bottom portion. Positive indexes generally indicate that the items “hang together” well and thus measure the same thing (which theoretically increases reliability). When some instructors find a near-zero or negative index for an item, they remove that item from their pool (for later semesters) or may possibly remove or downplay it in grading the current exam because this item may detract from reliability. Even less statistically-minded instructors may consider this exclusion strategy based on item information and general advice from test banks (e.g., Linden, 1998).

In this article, I suggest that this exclusion strategy in the use of discrimination indexes has potential pitfalls and may sometimes actually lower the validity of exams. This basic position is not new (e.g., see Cronbach, 1970; Henrysson, 1971; Kline, 1976), but this article offers additional justification and further discussion of the educational implications. An assumption in the use of such reliability statistics is that there is a single construct to be measured (e.g., see Cronbach, 1970; Funder, 1997; Kline, 1976), but this assumption may be generally untrue when it comes to course exams. A secondary assumption, which also seems questionable, is that the construct is stable (e.g., see Cronbach, 1970; Funder, 1997).

For example, an introduction to psychology midterm covers multiple topic areas, knowledge of which may not interrelate. Some instructors disagree, arguing for some form of a unitary intelligence construct, which may have some recent empirical support (see Kassin, 1998; Matarazzo, 1992), but this position also has numerous critics (e.g., Gardner, 1983, 1993; Sternberg, 1988). More important, Bloom (1976) and Kline (1976) identified students’ effort and ability as well as the ability of the instructor as distinct contributors to a course exam score, and at least students’ effort seems potentially unstable. Kline particularly distinguished one’s ability from the amount one has learned, noting that both are measured by a course exam but that they are “not necessarily highly correlated in any one individual” (p. 17). More generally, psychometricians have distinguished course exams from tests that measure single constructs, labeling course exams as attainment tests (Kline, 1976) and labeling various tests of single constructs as aptitude, ability, or personality tests (Cronbach, 1970; Funder, 1997; Kline, 1976).

If one treats an attainment test (or course exam) as if it were measuring a single construct, by requiring positive discrimination indexes, a primary pitfall is that content validity can be reduced. This cost would be incurred particularly if excluded items come from the same subarea (see Cronbach, 1970; Henrysson, 1971). However, Cronbach also argued that items with low indexes may reflect poorly written or ambiguous items or items that require knowledge beyond course content. In these cases, excluding the items seems justified because they do not allow a student to show his or her knowledge of the (predeclared) content. Thus, I am not recommending that instructors abandon the use of such statistics. Rather, I am suggesting that instructors not rely on the discrimination index alone in excluding any item, because the other potential causes of low discrimination indexes do not justify exclusion.

Two such causes mentioned by Cronbach (1970) are a topic’s level of difficulty and its low (or lack of) relation with other exam topics. Easy items also lead to lower indexes but should probably not be excluded for that reason alone (see Wesman, 1971). Although it seems relatively easy to see a systematic reason why “good students” (i.e., high exam scorers) answer an item correctly (leading to a positive discrimination index), the possibility of systematic reasons why “poor students” (i.e., low exam scorers) answer an item correctly (with as much or more frequency than “good students”) has received no attention. These reasons, which seem largely independent of the other causes, would also lead to lower index values but not justify excluding items.

Systematic Reasons Why “Poor Students” May Answer an Item Correctly

First, some students who are generally uninterested or less capable in a course may find one chapter or lecture particularly interesting or easy to understand. If the reason for this lack of interest or ability is shared by a group of students (e.g., belonging to a particular major), then they may also excel in the same subarea or find the same subarea interesting and study it more than other students. Thus, their overall exam scores may be relatively low, but not on the items testing that subarea, leading to lower discrimination indexes. For example, the math-related majors (or math-able students) who take a psychology course may do better on items regarding correlations or other statistical measures and worse on the more numerous psychology items, particularly as interest in math correlates negatively with interest in psychology (Stalder, 1997). Relatively, a subgroup of students may be entitled in some schools to organized tutoring outside of class, based on learning disability, ethnic background, or financial need. These students may tend to test poorly overall but may learn some difficult concepts very well because of the tutoring and the extra time devoted to those concepts. This possibility has been verified by a psychology professor who found negative discrimination indexes for a couple of items that seemed fairly worded and content valid (B. Bermant, personal communication, November, 1999). If items in such cases were removed or downplayed, then students who already face many learning obstacles would have one more.

Many other examples seem possible in my view (e.g., involving a subgroup of students with the same learning style), and these possibilities reflect the differences between the classroom and the laboratory. The dynamics of a course and its students and the variety of factors that can contribute to a student’s performance make it difficult to know why a discrimination index is low and thus what to do with that item. Admittedly, the larger the course’s enrollment, the less likely a small subgroup can mathematically influence discrimination indexes in the ways I described, but the size, number, and dynamics of these subgroups may be unknown in most cases.

Discrimination Power

Items with higher discrimination indexes tend to have greater discrimination power, which also has the potential to
reduce content validity (see Cronbach, 1970; Wesman, 1971). Items that are better discriminators may test more difficult concepts or may contain more difficult wording (e.g., double negatives, longer stems, or fine distinctions between choices; see Cronbach, 1970). In the case of wording difficulty, reading ability or logical reasoning ability may influence a student’s answer as much as or more than the student’s content knowledge and thus represents what Cronbach called an irrelevant difficulty. Thus, instructors who go to great lengths to write course exams that contain items with high discrimination indexes, or who restrict their exams to test bank items with high indexes, may inadvertently narrow their pool of items to one with a higher incidence of such difficulties. One end result is that higher grades may be awarded to students as much for their verbal and analytical skills as for their content knowledge. At an extreme, such outcomes could contribute to what some intelligence quotient critics have called an academic elitism (Lemann, 1999), in which it is largely those who “test well” (see Stanley, 1971) or who have high abilities in narrow domains (e.g., as tested by the Scholastic Aptitude Tests) who make it to the better universities or graduate schools. Assigning course grades does require some discrimination among students, but the range of difficulty in concepts and the differing levels of students’ studying and recall seem sufficient to ensure a reasonable range of grades.

In sum, near-zero or negative discrimination indexes can identify test items that are ambiguous or low in content validity, and these items deserve removal. In the absence of these factors, however, these items should probably be maintained. Otherwise, there are potential risks to content validity and to the subset of students who tend to answer such items correctly. In any case, the statistical assumption in this exclusion strategy, that a single construct is being measured, seems questionable for course exams (although factor analytic approaches might partially address this issue). Given in part these issues, I suggest a focus on content validity in the construction of course exams and encourage careful consideration of the potential causes of low discrimination indexes.

Conclusions

References


Use of In-Class Lab Groups to Enrich Independent Research Projects

Cynthia L. S. Pury
Clemson University

In this article, I suggest that the independent project required by most methodology courses can be profitably combined with a lab group model. Such in-class lab groups offer students the chance to explore an area of individual interest, while providing a forum for immediate peer feedback. In addition, students benefit from in-depth exposure to their group members’ projects.

Independent research projects are frequently assigned in research methodology courses (e.g., Chamberlain, 1986; Yoder, 1979). Such projects enhance student interest and allow instructors to evaluate students’ application of course material. The typical structure of independent projects, including midsemester design and individual consultation, demands much instructor time and can lead to student apprehension (Chamberlain, 1986).

Group projects offer an alternative (e.g., Chamberlain, 1986; Dunn, 1996). They teach teamwork and allow students to create more complex projects by division of labor. However, assigning individual grades for group projects is difficult, and social loafing can become problematic (Meyers, 1997).

Small, independent project lab groups can achieve many of the goals of independent and group projects, while avoiding many of their disadvantages. I have used such groups successfully for both introductory methodology and

Note
Send correspondence to Daniel R. Stalder, Department of Psychology, North Park University, 3225 West Foster Avenue, Chicago, IL 60625; e-mail: dstalder@northpark.edu.
I describe a systematic approach to study tips whereby I provide students with a handout detailing several tips designed to promote self-assessment and comprehension. A total of 114 introductory psychology students rated the helpfulness of, and the degree to which they used, 6 tips. They rated the lecture-notes tip the most helpful; tips regarding a course-performance chart, elaborative encoding, and question and answer as intermediate in helpfulness; and self-help quizzes and study-group tips, least helpful. There was a small but significant correlation between the students’ helpfulness ratings of the tips overall and their course performance to that point. Mean degree-of-use ratings for each tip showed the same order as mean helpfulness ratings. Overall degree-of-use ratings were correlated with overall helpfulness ratings, but not with course performance.

The amount of research on study skills has been prodigious. For instance, Hattie, Biggs, and Purdie (1996) found 1,415 journal articles reporting research on study skills published between 1982 and 1992. Frequently, such research has focused on self-assessment (e.g., Prohaska, 1994; Sjostrom & Marks, 1994) or comprehension (e.g., Bartling, 1988; Hart & Speece, 1998).

Despite this abundance of research, instructors often face the problem of how to help students study more effectively. Books (e.g., Heffernan, 1997; Hettich, 1998; Walter & Selbert, 1987) or courses (e.g., Petrie & Helmcamp, 1998) on study skills usually target a level of education (e.g., high school or undergraduate) or field of study (e.g., psychology) rather than particular courses. Thus, a given instructor’s students may not have had such help, or it may have been inappropriate for the instructor’s course and teaching methods. The current consensus seems to be that study skills are best taught when linked to a particular content area (Garner, 1990; Hattie et al., 1996).

This article reports my attempt to help introductory psychology students with study skills. Class time focused mainly on the content of 14 subtopics in psychology. I decided to distribute my own study-tips handout, requiring only about 15 min to explain to the students during the first class. The handout included those tips that I deemed most relevant to the course and that were linked in the literature to good academic performance. All the tips required some degree of effort and initiative from students. Students are unlikely to use, or benefit from, tips that they consider unhelpful. Thus, I evaluated both their helpfulness and their degree-of-use ratings of the tips, hoping that such information would be useful to instructors wishing to provide their students with such tips.

In Spring 1999, 114 students enrolled in my two sections of Introductory Psychology completed an evaluation of the study tips.

Study-Tips Handout

During the first class, I distributed and explained to students a handout that included six study tips. The self-help quizzes, course-performance chart, and question-and-answer tips were designed to promote self-assessment, whereas the lecture notes, study groups, and elaborative-encoding tips were designed to promote comprehension.

Self-help quizzes. I advised students to take, after studying each chapter and the accompanying lecture notes, two self-help quizzes, one based on odd-numbered (SHQ1) and even-numbered (SHQ2) sample multiple-choice questions in the study guide to the textbook (Carlson & Buskist, 1997). They would take SHQ1 as if it was a real quiz, score it, calculate their percentage correct, and compare it to the one required for the grade that they were aiming for. After further study, they would take SHQ2 in a similar manner. Thus, they could assess their improvement, readiness to take the real quiz, and need for any further study. This tip is consistent with the finding that feedback from practice tests improves performance on target tests (Balch, 1998; Glenberg, Sanocki, Epstein, & Morris, 1987).

Course-performance chart. This chart, attached to the study-tips handout, provided a format for students to record each quiz or test score, calculate their percentage correct, and estimate their grade throughout the course. I suggested that students use their scores to adjust their study, concentrating on whatever techniques they used for quizzes or tests on which they scored well. This tip is consistent with the principles of metacognition and self-assessment (Flavell, 1979; Van Ede, 1996).

Question and answer. This tip involved a student answering questions that another person (e.g., classmate, roommate, etc.) asked while referring to the student’s text or lecture notes. Students should be able to answer in their own words, without looking at the text or notes. If they were unsure of their accuracy, they could check their written materials afterward. Although I know of no studies directly addressed to whether answering such questions improves performance, writing (and presumably knowing the answer to) test-relevant questions improves performance on a test (Foos, 1989).
Lecture notes. I advised students not to try to record every word I said. Instead, they should listen to each point in the lecture and, immediately after understanding the point, write it down in their own style. They should also review notes as soon as possible after each lecture, trying to reconstruct the lecture to remember it better. This tip is consistent with findings that the amount and kind of information included in lecture notes are related to test performance (Baker & Lombardi, 1985).

Study groups. I recommended that students, after studying individually, study further in small groups. A student who understands a certain concept could better explain that concept to a student having difficulty with it. Other researchers and teachers have also recommended this technique (e.g., Ender, 1985; Hettich, 1998).

Elaborative encoding. I suggested that students elaborate while studying. To illustrate, I used the various components of audition (e.g., ear drum, hair cells, auditory nerve, etc.). I advised students not to try to memorize the definition of each component verbatim. Instead, they might describe in their own words how a sound is processed and then finally, bringing in all the components and explaining what role each plays. Students' use of elaborative encoding is linked to their academic success (e.g., Chi, Bassok, Lewis, & Reimann, 1989; Gadszella, 1995).

Study Tips Evaluation

During the last class of the semester, students rated each study tip both on how much it "helped [their] performance in the course" (helpfulness) and on "how much [they] used" it (degree of use). They did so by making a vertical mark on each of six 10-cm horizontal-line scales, ranging from 0 (not at all) at the left and 10 (very much) at the right. Each scale was labeled on top with just the name of the tip. However, students could refer to the original study-tips handout if needed. In addition, they rated—on a similar scale—how much the tips helped them “overall.” They completed all seven helpfulness ratings first, and then all seven degree-of-use ratings.

Results

Each rating varied from 0 (not helpful at all or not used at all) to 10 (very much helpful or used very much). Table 1 shows the mean and standard deviation of the ratings for each study tip as well as for the tips overall. The mean helpfulness ratings ranged from 2.8 to 6.0, and the mean overall rating was 5.0 (SD = 2.3). The mean degree-of-use ratings ranged from 2.6 to 6.3, and the mean overall rating was 4.0 (SD = 2.3). Note the identical orders of the mean helpfulness and degree-of-use ratings of the individual tips. In addition, overall helpfulness and overall degree-of-use ratings were correlated, r(113) = .42, p < .001.

Helpfulness Ratings

A one-way, within-participants ANOVA on the six individual-tip ratings served to assess the relative helpfulness of each tip. The study-tips effect was significant, F(5, 565) = 20.25, p < .001, with an effect size of η2 = .15. Bonferroni pair-wise comparisons on the ordered means of the helpfulness ratings divided the tips into three levels: most helpful (lecture notes), immediately helpful (course-performance chart, elaborative encoding, and question and answer), and least helpful (self-help quizzes and study groups).

The correlation between overall helpfulness ratings and course performance (as measured by all-but-final point totals) was small but significant, r(113) = .26, p < .01. Apparently there was some objectivity to these ratings. Possibly the study tips increased performance for students who used them more. Alternately, students who performed better may have interpreted the study tips retrospectively as having helped them more. However, helpfulness ratings of individual study tips did not correlate with course performance, ps > .05, except for ratings of the course-performance chart, r(113) = .19, p < .05. Ratings of lecture notes and of elaborative encoding were marginally correlated with course performance, ps < .10.

Degree-of-Use Ratings

To assess the relative degree of use of each tip, I performed a one-way, within-participants ANOVA on the individual-tip ratings. The study-tips effect was again significant, F(5, 565) = 37.42, p < .001, with an effect size of η2 = .25. Bonferroni pair-wise comparisons on the mean ratings of the individual tips divided the tips into three levels: most used (lecture notes and course-performance chart), intermediate used (elaborative encoding and question and answer), and least used (self-help quizzes and study groups).

There was no correlation between overall degree-of-use ratings and course performance, r(113) = -.05, p > .10, suggesting that the objectivity of those ratings is questionable. Yet there was a significant correlation between course performance and degree-of-use ratings for elaborative encoding, r(113) = .19, p < .05, although none of the compa-

<table>
<thead>
<tr>
<th>Study Tip</th>
<th>Helpfulness Rating</th>
<th>Degree-of-Use Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture notes</td>
<td>6.0b</td>
<td>6.3b</td>
</tr>
<tr>
<td>Course-performance chart</td>
<td>4.9b</td>
<td>5.9b</td>
</tr>
<tr>
<td>Elaborative encoding</td>
<td>4.8b</td>
<td>4.8b</td>
</tr>
<tr>
<td>Question and answer</td>
<td>4.6b</td>
<td>4.2b</td>
</tr>
<tr>
<td>Self-help quizzes</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Study groups</td>
<td>2.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Overall study tips</td>
<td>5.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Note. For each entry, n = 114.

b Each student rated all study tips for helpfulness and degree of use, using an 11-point scale ranging from 0 (not at all) to 10 (very much). The mean ratings of the six individual tips (not counting overall study tips) appear in the order of descending helpfulness or degree of use.

de Each of these footnotes indicates tips that do not differ, as determined by five pair-wise Bonferroni comparisons performed separately for each type of rating. For significant differences between levels of tips, ps < .025.
rable correlations were significant for the other individual study tips, ps > .10.

Discussion

Students’ mean overall helpfulness rating of the study tips was only midway along the helpfulness scale, reflecting only moderate enthusiasm about the tips. Their mean overall degree-of-use rating was a little below midway along the scale, reflecting their honesty but also suggesting that their motivation to use the tips was not overwhelming. Perhaps if I had spent more class time explaining and promoting the tips, the overall degree-of-use ratings would have been higher.

Although all the tips stemmed from research on effective study skills, the students showed clear differences in how they rated tips both on helpfulness and on degree of use. These differences might reflect actual differences in the effectiveness of the tips or might reflect differences in the effort required by the tips. For instance, self-help quizzes and study groups may require the most time and organization. For that reason, students may rate these tips low. Another interpretation is that tips involving more overlap with test questions may be rated more helpful. For example, many of my quiz and test questions came from lecture material but fewer questions may have come from self-help quizzes or questions asked in study groups. Finally, students tended to rate the tips that they used more as being more helpful. The identical orders of the mean helpfulness and degree-of-use ratings for each tip, and the significant correlation between the overall helpfulness and degree-of-use ratings for each tip, and the overall degree-of-use ratings, support such a conclusion.

Whatever the students’ reasons for their ratings, an instructor’s knowledge of their differential reactions to study tips is useful in recommending such tips to students.

References


Notes

1. The research reported in this article was supported by a faculty development grant from Pennsylvania State University, Altoona.

2. Send correspondence to William R. Balch, Psychology Department, Pennsylvania State University, Altoona, 3000 Ivyside Park, Altoona, PA 16601–3760; e-mail: wrb3@psu.edu.

Instructor Evaluations of Introductory Psychology Teaching Techniques

Lee I. McCann
Baron Perlman
Tanya L. De Both
University of Wisconsin Oshkosh

We asked experienced introductory psychology instructors to rate the perceived effectiveness of 22 teaching techniques to improve student performance. These teachers rated more in-class examples and activities, writing assignments, emphasizing core concepts in lectures and exams, and providing copies of lecture outlines as most effective. There was no significant correlation between a method’s frequency of use and its perceived effect on student learning. Instructors can use perceived effectiveness information to assist in selecting methods to improve their teaching.

Psychology faculty have a well-documented history of efforts to increase student interest and performance, and research on the effect of specific pedagogical techniques is common. Evaluation of such methods has typically compared a traditional class with another taught using a specific
Student Perspectives on Grade Changes From Test to Test

Baron Perlman  
Lee I. McCann  
University of Wisconsin–Oshkosh

We surveyed undergraduate psychology students whose exam grades had gone up or down a minimum of 1 full letter grade from 1 exam to the next. We asked about their perceptions of how their studying and preparation differed for each exam. These data may be useful for students and may assist faculty in advising about better exam preparation and performance.

When students want to improve their exam performance what should faculty tell them? We found no information about influences on student test performance in popular teaching books (e.g., Davis, 1993; McKeachie, 1999) and few empirical studies that investigated the relationship between student classroom behavior (e.g., lecture notes; Baker & Lombardi, 1985) or external influences on students (e.g., life events; De Meuse, 1985) and exam performance.

Equally or more important, when faculty give such advice, how do they maximize the probability that students will follow through on their suggestions? Letters to Successors (Brookfield, 1995) is a technique in which, at the end of a semester, students list what it is essential to know and do to survive or succeed in the course, and the instructor presents this information to students the next time the course is taught. This student perspective may be more influential than faculty advice. We used a variation of this method to seek student perspectives on the specific issue of how their studying and preparation influence exam performance.

Notes

1. I thank B. J. Paschal for his contribution in conducting this study and Sharon E. Paulson for her guidance in editing. I thank Jack Snowman for his permission to use his article (see the Reference section).
2. Send correspondence and requests for the quizzes and survey used in this study to Gregory J. Marchant, Department of Educational Psychology, Ball State University, Muncie, IN 47306–0595; e-mail gmarchant@bsu.edu.

Method

Participants

Eleven of 12 full-time faculty members in a regional public university sampled undergraduate students whose exam score went up or down substantively (at least one full letter grade) in their psychology courses (n = 26) during Fall 1998 and Spring 1999 semesters. Participants were 340 students (165 freshmen and 175 sophomores, juniors, and seniors; 218 women and 122 men). Most students (n = 246) were enrolled in General Psychology, with 94 in sophomore and upper level courses.

Questionnaire

Students responded to a questionnaire (based on our own and departmental colleagues’ teaching experiences) that listed 19 behaviors that might improve and 17 that might negatively affect test performance; students could add “other” responses. We asked for students’ gender, year in college, and course level (first, second, third, or fourth year). The questionnaire took 5 min to complete.

Procedure

Faculty identified students whose exam scores had gone up or down at least one full letter grade (e.g., not a C+ to a B–). Faculty had the freedom to decide whether that definition was too lenient (i.e., a student’s full letter-grade change was not a substantive one based on the faculty member’s grading criteria). For each student with a substantive grade change, faculty highlighted a column on the questionnaire to indicate whether the student “did better” or “did worse,” and distributed questionnaires before or after class. Students returned the questionnaire in a sealed envelope to the instructor or department secretary (name and office number listed). Students checked all items they thought contributed to their exam grade as compared to the previous test and identified the one item they believed most important in explaining their grade change. Faculty instructed students to participate only once, and one item asked students to check if they had filled out the questionnaire previously in this or another course.

After the second exam faculty told all students:

I am assisting in a study to determine what factors contribute to a student doing better or worse on exams. We hope the information you provide will help us to better advise students in the future. This study is optional; you do not have to participate. All data are confidential and your responses cannot be attached to your name. There is potential benefit to participating. In reading the list of behaviors related to test grades, you may get ideas on how to do better in this course. By turning in this brief questionnaire, you agree to participate in the research.
We could not sample students in courses with only two exams because students typically took the second exam on the last day of class. We used courses that had three \((n = 7, 27\%)\) or four exams \((n = 19, 73\%)\), comparing students’ performance from the first to second or second to third test. Exams varied from all multiple choice, all essay, etc.

**Results**

Of 550 questionnaires distributed, 419 were returned \((76\%)\). We discarded 79 \((student filled one out previously, n = 43; missing demographic information, n = 24; students checked items in both the “did better” and “did worse” columns, n = 8; or students who did better checked “did worse” items or vice versa, n = 4)\. The usable response rate was 62%; we do not have response rate data separated by better or worse exam grades. Students listed significantly more reasons for test grades going up \((< .01\)) than down \((< .01\) or students who did better checked “most important” items. We used a Bonferroni correction \((Hays, 1988)\) to minimize Type I error and a significance level of .05. None of the 40 chi-square analyses were significant, indicating that gender, course level, and year in school were not related to perceived reasons for doing better. We present overall data in Table 1.

We expected more data for students whose grades had risen than fallen because many students improve as a course progresses, and almost 70% of the usable sample did involve exam grades going up \((n = 236; 81 men and 155 women; 124 freshmen and 112 sophomores, juniors, or seniors; 184 enrolled in general psychology, 52 in second-, third-, and fourth-year courses). We compared men and women, first-year students with all others, and general psychology with students in other courses for the (a) 10 most frequently checked behaviors for test scores rising and (b) most frequently checked “most important” items. We used a Bonferroni correction \((Hays, 1988)\) to minimize Type I error and a significance level of .05. None of the 40 chi-square analyses were significant, indicating that gender, course level, and year in school were not related to perceived reasons for doing better. We present overall data in Table 1.

Students identified studying more hours \((n = 38, 16\%)\) as the single most important behavior in improved exam performance, followed by reading more carefully \((n = 37, 16\%)\), using the study guide more \((n = 24, 10\%)\), and reading material more times \((n = 14, 6\%)\). Students rated 19 other behaviors as a “most important” perceived influence on raised exam scores. Just over 30% of the usable sample \((n = 104)\) returned questionnaires for falling exam grades \((41 men and 63 women; 43 freshman and 61 sophomores, juniors, or seniors; 62 enrolled in General Psychology and 42 in second-, third-, and fourth-year courses). We again used a Bonferroni correction to minimize Type I error. None of 33 chi-square analyses for the 10 most frequently checked behaviors related to test scores falling or for most frequently checked “most important” items were significant, \(p < .05\), indicating that gender, course level, and year in school were not related to perceived reasons for doing worse. Lowered grade data appear in Table 2.

The single most important behavior related to a poorer test grade was studying fewer hours \((n = 22, 21\%)\), followed

### Table 1. Contributions to Improved Examination Grades

<table>
<thead>
<tr>
<th>Reason</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied more hours</td>
<td>156</td>
<td>66</td>
</tr>
<tr>
<td>Read the book/text/readings more carefully</td>
<td>154</td>
<td>66</td>
</tr>
<tr>
<td>Studied in a quiet environment, few interruptions</td>
<td>117</td>
<td>50</td>
</tr>
<tr>
<td>Used the study guide more</td>
<td>75</td>
<td>32</td>
</tr>
<tr>
<td>Read the text/readings more times</td>
<td>72</td>
<td>31</td>
</tr>
<tr>
<td>Took better lecture notes</td>
<td>54</td>
<td>23</td>
</tr>
<tr>
<td>Took notes on readings, had not before</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td>Cut back on time socializing/partying</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>Got lecture notes I missed from someone</td>
<td>31</td>
<td>13</td>
</tr>
<tr>
<td>Made flashcards</td>
<td>27</td>
<td>11</td>
</tr>
<tr>
<td>Studied with someone, had not before</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Worked fewer hours at part- or full-time job</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Easier/more interesting material, understood the material better</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Talked with the instructor, clarified how and what to study</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>Came to class more regularly</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>Family/personal crisis or problem resolved</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Different approach to studying not defined</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Reviewed the last exam</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Lower stress level, more sleep</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Worked with a tutor</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Studied lecture notes more carefully</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Began using university support services e.g., reading center, writing center</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Paid more attention in class</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Bought the text, had not had it previously</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Other classes did not take time away</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

\(n = 236\). \(^{n}\)Added by students under the “other” category.

### Table 2. Contributions to Worse Examination Grades

<table>
<thead>
<tr>
<th>Reason</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied fewer hours</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>Read the book/text/readings less carefully</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>Read the text/readings fewer times</td>
<td>43</td>
<td>41</td>
</tr>
<tr>
<td>Family/personal crisis or problem interfered</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Missed more classes</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Missing lecture notes</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Studied in a noisy environment, many interruptions</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Worked more hours at part- or full-time job</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Spent more time socializing/partying</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>Used the study guide less</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Took worse lecture notes</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Did not make flashcards this time</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Stopped taking notes on readings</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Took notes on reading, did not help and took time away from more productive studying</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Other tests/assignments due at the same time</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Studied with someone, actually was a hindrance</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>No book, did not read the book</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Less straightforward test questions, more difficult material</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Do not know, studied the same</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Stopped studying with someone</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Missed class due to extracurricular activity</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Took different kind of notes on reading</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Could not finish test on time</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Distracted in class</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Studied wrong information, did not know what to focus on</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Other: lack of sleep, did not feel like studying, could not stay on task, too much TV</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

\(n = 104\). \(^{n}\)Added by students under the “other” category.
by personal problems or crises (n = 10, 10%). Students rated 17 other behaviors as a “most important” perceived influence on falling test performance.

Discussion

In most cases, the data complement accepted faculty wisdom on this topic. Some students perceive studying harder as related to rising test grades and less studying to lower grades. However, more is going on.

We were surprised by some of the results. Students did not rate some behaviors faculty might expect to affect test grades as doing so (e.g., regularly attending class, easier or interesting course material, tests occurring at the same time). Also, students’ perceptions did not differ based on year in school, gender, or course level for either rising or falling grades. These findings suggest that the data in Tables 1 and 2 provide good overall lists for any student of what “to do” and “not to do” when studying for exams. Another surprise was that the data did not support student helplessness and lack of control over exam performance as explanations for doing better or worse. We found few comments such as “I don’t know what happened” or “I studied the same way, my grade just went down.”

Sharing these student perceptions may help all students maximize exam performance, and they may be useful when faculty advise individual students. Because many behaviors perceived as related to exam performance seem independent of any specific course or teacher behavior (e.g., life crises or hours worked), faculty should emphasize to students that problems with exam performance, or strategies to improve them, may have effects beyond the current course, influencing academic performance throughout a student’s educational career.

References


Notes

1. This article is based on a poster presented at the 23rd annual National Institute on the Teaching of Psychology, January 2001, St. Petersburg Beach, FL.
2. We gratefully acknowledge the assistance of our students and departmental colleagues.
3. Send correspondence and requests of the questionnaire to Baron Perlman, Department of Psychology, University of Wisconsin, Oshkosh, WI 54901; e-mail: perlman@uwosh.edu.

Student Versus Faculty Perceptions of Missing Class

Merry J. Sleigh
Darren R. Ritzer
George Mason University
Michael B. Casey
College of Wooster

We examined student and faculty perceptions of class attendance and of professor–student interactions regarding missed classes. Students found more reasons for missing class acceptable and found it more important to explain an absence than did faculty. Faculty predicted that the interest level of the material would be the most important factor in determining class attendance; however, students reported that the amount of class material on the exam was the most important factor. Grade point average, number of credits accumulated, and gender of the student related to student perceptions. Number of years teaching or gender of the faculty did not significantly relate to faculty perceptions.

Research has indicated that students and faculty often perceive the same situation in different ways. For example, faculty and students differ in their perceptions of the academic honesty of undergraduates (Smith, Nolan, & Dai, 1998), the meaning of student grades (Goulden & Griffin, 1997), and the value placed on service learning (Rowe & Chapman, 1999). One relatively unexplored area in which students and faculty may have different views is reasons for attending or not attending class. These perspectives are important for educators to examine because class attendance relates to course grades (Buckalew, Daly, & Coffield, 1986; Simpson & Nist, 1992; Van Blerkom, 1992), student retention (Rotter, 1988), and the development of academic skills (Terenzini, Theophiles, & Lorang, 1984). The benefits achieved from class attendance are even more pronounced in younger college students (Brocato, 1989).

In this study we examined the ways that students perceive class attendance and reasons for missing class and compared those perceptions to faculty perceptions. This examination has important implications for student–faculty interactions and classroom management.

Method

Participants

Participants included 231 undergraduates (89 men and 142 women) from a large Midatlantic university. Thirty percent were freshmen, 26% were sophomores, 22% were juniors, and 22% were seniors. A total of 38% had declared a psychology major; the remainder were undeclared or majoring in other subjects. The mean grade point average reported was 3.05 (SD = .52, range 1 to 4).

Faculty participants included 22 full-time faculty (16 men and 6 women) in the psychology department at the same uni-
How Do Students Really Study (and Does It Matter?)

Regan A. R. Gurung

Human Development and Psychology
University of Wisconsin, Green Bay

There are many different ways to study but not all methods may enhance learning. Although there is a sizeable literature on how students should study (Al-Hilawani & Sartawi, 1997; Fleming, 2002; Hattie, Biggs, & Purdie, 1996), not as much is known regarding how students actually do study. I assessed how students actually study and tested whether certain study habits were more conducive to learning than others.

Study skills can be divided into four main categories: repetition-based (e.g., flashcards and mnemonic devices such as “CANOE” for the Big 5 personality traits), cognitive-based (e.g., studying with a friend, group work), procedural (e.g., time management, organization, scheduling study routines), and metacognitive (e.g., taking quizzes to test self-knowledge; for more details, see Gettinger & Seibert, 2002). Empirical tests comparing these different methods are equivocal.

Some research suggests that the types of study techniques that a student uses affect exam performance (Bol, Warkentin, Nunnery, & O’Connell, 1999). Other research suggests that there is no one style that is useful for everyone and that a repertoire of techniques is best (Hadwin & Winne, 1996; Nist, Simpson, Olejnik, & Mealey, 1991). For example, repetition and rehearsal, which require minimal amount of processing, may be useful only in remembering small amounts of information (Gettinger & Seibert, 2002). Memorizing facts and definitions does not correlate with students’ exam scores, but procedural and organizational-based skills, metacognitive-based skills, and skills that increase elaboration show positive correlations with test scores (Carney & Levin, 1998; Chen & Daehlter, 2000; Elliott, McGregor, & Gable, 1999; Motes & Wiegmann, 1999). Dickinson and O’Connell (1990) also showed that time spent organizing course material (e.g., taking notes on the textbook) related to test scores, whereas actual hours spent studying did not.

The existing literature does not include a comprehensive assessment of a wide variety of studying techniques, and it does not provide studies that both assess techniques and measure learning outcomes. Furthermore, students are often unaware that some of their habits, such as having music on while studying, may hurt their learning. This study provides a rich view of what students do by collectively assessing different behaviors. Consistent with the disparate literatures, I hypothesized that techniques aiding elaboration (e.g., using examples, mnemonics) and metacognition (e.g., self-testing) would predict higher exam scores, whereas those behaviors reducing elaboration (e.g., listening to music) would predict lower exam scores.

Notes
1. We thank Alejandro Lazarte and Jared Keeley for their assistance with statistical analysis.
2. Send correspondence to William Buskist, Psychology Department, Auburn University, Auburn, AL 36849–5214; e-mail: buskifw@auburn.edu.

Method

Participants

Two hundred and twenty-nine students (169 women and 60 men) from a midsized midwestern university in two sections of my introductory psychology class participated in this study (participation was voluntary). The mean age was 19.26 (SD = 3.91). The majority of the students were freshmen (82%); the remainder were sophomores (7%), juniors (4%), and seniors (7%). The mean ACT score was 22 (range 10 to 31). I combined the data from both sections as exam grades were similar.
A questionnaire assessed study methods, distractions, and confidence with the material. I based items on previous research (Wade, Trathen, & Schraw, 1990; Winne & Jamieson-Noel, 2002) and feedback from small student focus groups (questionnaire available on request). I asked students which of 11 study methods they used (i.e., memorizing definitions, reading the text, reviewing figures, reviewing highlighted material in the text, testing self-knowledge, rewriting notes, taking notes on the text, mnemonics, studying with friends, reading the notes, rewriting notes) and the extent to which they used them on a 5-point scale ranging from 1 (never) to 5 (all the time). I also measured distractions ("Do you have music or the television on when studying? Do you have roommates, family, or friends around when studying?" and "Do you respond to instant messaging or e-mail while studying?"); the total hours students studied for the exam; the number of days in advance that students started studying; how often they reviewed material before and after a class; and how well they believed they knew the material, understood the material, and how confident they were of their understanding of the material.

Procedure

I added the survey to the end of the last of four exams. After answering 65 multiple-choice questions, participants read instructions stating that the remaining questions on the exam sheet would assess their study habits. I told students that participation was voluntary and that the answers to the questions would not affect their class grades or exam scores.

Results

The majority of students reported studying between 4 to 6 hr for the final (45%). The rest studied between 1 to 3 hr (31%) and 7 to 9 hr (19%). A small number of students reported studying over 10 hr (5%). The frequency and duration for use of the 11 study techniques assessed in this study appear in Table 1.

The frequency of technique use and the duration of technique use were correlated with scores on students’ final exam. Partial correlations controlled for student ability (using ACT scores; zero-order correlations available on request). The more students memorized notes, $r(227) = .28, p < .001$; made up examples $r(227) = .20, p < .001$; read the book $r(227) = .21, p < .01$; read their notes, $r(227) = .18, p < .05$; used mnemonics, $r(227) = .15, p < .05$; and tested their knowledge, $r(227) = .28, p < .001$; the higher were their exam scores. No other techniques (i.e., frequency of use) significantly correlated with exam score.

In contrast to the significant correlations with frequency of use described previously, only the amount of time spent memorizing was significantly related to exam scores, $r(227) = .15, p < .05$. The global number of hours studied did relate to exam scores, $r(228) = .16, p < .05$.

All the distracters and not attending class negatively correlated with exam grades. Students who had music on, $r(226) = –.18, p < .01$; the television on, $r(226) = –.21, p < .01$; responded to e-mail, $r(226) = –.16, p < .05$; or who had friends around, $r(226) = –.13, p < .05$; when studying performed worse on the exam. Students who missed class also scored lower on the exam, $r(226) = –.27, p < .001$.

Table 1. Frequency and Duration of Use of the Main Study Techniques

<table>
<thead>
<tr>
<th>Study technique*</th>
<th>Hours Spent (%)</th>
<th>M</th>
<th>0</th>
<th>1</th>
<th>1 to 2</th>
<th>2 to 3</th>
<th>&gt; 3</th>
<th>How Often</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read your notes</td>
<td></td>
<td>1</td>
<td>18</td>
<td>32</td>
<td>26</td>
<td>16</td>
<td>4.01</td>
<td></td>
</tr>
<tr>
<td>Read the text</td>
<td></td>
<td>4</td>
<td>23</td>
<td>34</td>
<td>20</td>
<td>11</td>
<td>3.37</td>
<td></td>
</tr>
<tr>
<td>Think of mnemonic devices (e.g., “CANOE” for personality traits)</td>
<td></td>
<td>13</td>
<td>41</td>
<td>23</td>
<td>10</td>
<td>3</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>Rewrite notes and/or skim notes</td>
<td></td>
<td>10</td>
<td>28</td>
<td>32</td>
<td>14</td>
<td>8</td>
<td>3.25</td>
<td></td>
</tr>
<tr>
<td>Review highlighted information from text</td>
<td></td>
<td>8</td>
<td>34</td>
<td>35</td>
<td>11</td>
<td>2</td>
<td>3.15</td>
<td></td>
</tr>
<tr>
<td>Memorize definitions through repetition (e.g., flashcards)</td>
<td></td>
<td>9</td>
<td>36</td>
<td>31</td>
<td>13</td>
<td>4</td>
<td>3.11</td>
<td></td>
</tr>
<tr>
<td>Review figures and tables in text</td>
<td></td>
<td>8</td>
<td>51</td>
<td>24</td>
<td>7</td>
<td>1</td>
<td>2.96</td>
<td></td>
</tr>
<tr>
<td>Make up examples to understand material/incorporate into everyday life</td>
<td></td>
<td>16</td>
<td>43</td>
<td>23</td>
<td>9</td>
<td>2</td>
<td>2.89</td>
<td></td>
</tr>
<tr>
<td>Use concept checks, chapter-end questions to test knowledge</td>
<td></td>
<td>23</td>
<td>42</td>
<td>19</td>
<td>6</td>
<td>2</td>
<td>2.62</td>
<td></td>
</tr>
<tr>
<td>Take notes from the book</td>
<td></td>
<td>39</td>
<td>38</td>
<td>14</td>
<td>8</td>
<td>1</td>
<td>2.18</td>
<td></td>
</tr>
<tr>
<td>Study with a friend</td>
<td></td>
<td>43</td>
<td>29</td>
<td>18</td>
<td>8</td>
<td>2</td>
<td>2.07</td>
<td></td>
</tr>
</tbody>
</table>

Distracters

| Have the television on | 4.00 |
| Have music on | 3.78 |
| Have roommates/friends/family around | 3.02 |
| Respond to instant messaging/e-mail on the Internet | 3.75 |

Self-reports level

| Knowledge | 3.43 |
| Understanding | 3.66 |
| Confidence | 3.44 |

Note. Hours spent represents the percentage of the class who reported using each option. How often each technique was used/or each distraction was present was measured on a 5-point scale ranging from 1 (never) to 5 (all the time). Knowledge, understanding, and confidence was measured on a 5-point scale ranging from 1 (not at all) to 5 (extremely).

*Listed in order of use.
Discussion

The results of this study provide a detailed picture of what students spent time on and how effective the different methods were. Not all techniques were effective—the most effective techniques were often not the ones used the most. For example, although the three most frequently used techniques (reading notes and the text, using mnemonics) correlated significantly with exam scores, one of the strongest predictors of exam scores, testing knowledge, was one of the least used techniques. Other techniques commonly used by students (rewriting notes, reviewing highlighted material and figures and tables in the text) did not relate to exam scores. Perhaps most important, the number of hours studied was only weakly associated with exam score. This finding suggests that how students study may be even more important than how long they study and provides a strong rationale for the use of this measure.

The effectiveness of many common study suggestions did not receive empirical support. Results such as these compel a closer look at the recommendations instructors make to their students. Instructors often provide study tips and urge students to use specific techniques, but the correlation between student grades and technique usage is not always significant and is often low. For example, Balch (2001) provided students with six tips (e.g., specific ways to take lecture notes and self-help quizzes) but, except for elaborative encoding, reported use and course grades were not significantly correlated.

The results provide strong empirical evidence of what students should not do. Students who skip class, listen to music, watch television, or use the Internet while studying performed worse on the exam. Although the data do not test the causality of the association between distractions and exam scores, making such data available to students on the first day of class may help them better design their study habits.

This study provides an easy method for individual instructors to assess how their students are preparing for exams. These findings may vary for instructors who use essay exams or different textbooks, but an instructor can modify this method to assess different levels of classes and for different types of exams. The results of this study and self-collected data with this tool will prepare instructors to advise students on how best to study to do well.

Limitations

Finals are a particularly stressful time of the semester, and studying during the last week of class may not be representative of how students study in general. The fact that the exam was not cumulative (similar to midterm exams) somewhat lessens the problem with this limitation. Having the students complete the assessment as part of the exam (and hence be identifiable) raises the potential for impression management and could contaminate responding. Finally, I made the assumption that exam scores equate to learning. It is possible that even the study techniques that did not significantly relate to exam scores did enhance learning, but this learning was not captured by my exam.

Although students use a variety of study techniques, they are not all effective. Furthermore, students are not using some useful techniques enough. How students prepare for tests can be a crucial element in their achievement. Because certain study techniques are more beneficial than others, instructors should help students more effectively prepare for exams by informing students about the techniques and modifying ways to best help students use the techniques. How students study does actually seem to matter.

References


Notes

1. The University of Wisconsin, Green Bay, Department of Human Development funded this research.

2. A portion of this study was presented at the American Psychological Society Meeting in Chicago in 2004.

3. I thank Heidi Rose and Jessica Peterson for their help collecting, analyzing, and discussing the data and Heather Bloch, Sarah Brill, and Ilene Noppe for their helpful suggestions.

4. Send correspondence to Regan A. R. Gurung, Department of Human Development and Psychology, University of Wisconsin, Green Bay, 2420 Nicolet Drive, MAC C318, Green Bay, WI 54311; e-mail: gurungr@uwgb.edu.
sons learned from the examples (i.e., plagiarism occurs in the real world) when creating their new paragraphs. Adopting a proactive approach to eliminate plagiarism is important because students who are unclear about plagiarism may assume that they are sufficiently knowledgeable and, consequently, may not seek greater understanding. Likewise, instructors who assume that students know how to avoid plagiarism may miss an important opportunity to give students the skills to avoid the consequences of academic dishonesty. The implications of this study are twofold. First, placing a nonspecific directive to “avoid plagiarism” on a syllabus or making a similarly vague statement in class is not as effective as providing students with performance feedback or examples of plagiarized passages. Second, it is neither difficult nor time consuming to effect a change in students’ ability to detect and avoid plagiarism. In one class period, students learning to write an American Psychological Association style research report can study examples of plagiarism, take a test on their knowledge of it, and receive feedback on their ability to avoid it. Although the results indicate that our experimental manipulations may be useful, some cautions are in order. We did not test whether these effects would extend to assignments later in the semester or when a grade is at stake. Despite these limitations, our findings suggest that instructors should not feel helpless in preventing plagiarism. Giving examples or feedback on paraphrasing attempts can have a positive effect on students’ knowledge of plagiarism and their ability to avoid it.

References


Notes

1. These data were presented at the annual meeting of the Eastern Psychological Association, Providence, RI, April 1999.

2. We thank Miguel Roig for generously sharing his materials and Claire Klinedinst and Susan Campbell for allowing us to use their class time. We also thank Randolph A. Smith, Linda M. Noble, Andy Leynes, and three anonymous reviewers for their comments on an earlier draft of this manuscript.

3. Send correspondence to Joshua D. Landau at Department of Behavioral Sciences, York College of Pennsylvania, York, PA 17405–7199; e-mail: jlandau@ycp.edu.

---

**Improving Students’ Exam Performance by Introducing Study Strategies and Goal Setting**

Victoria Manion Fleming  
*Department of Educational Psychology*  
*Miami University*

This study addresses whether learning strategies would improve students’ exam performance. Students in 2 sections (N = 65) of Introductory Psychology participated. I introduced students in the experimental section to learning strategies. Students set individual learning goals and recorded their learning related behavior during the first 2 units. Students in the control condition engaged in nonacademic tasks. All students experienced a lesson on learning at the end of Unit 2. First-year students in the control condition ob-
Can instructors teach students learning strategies without interfering with the curriculum, and will that instruction result in more learning? Some have argued that self-regulated learners have the best chance for educational success (McCombs & Marzano, 1990). To be self-regulated learners, students need knowledge about learning strategies (Weinstein, Meyer, & Van Mater Stone, 1994; Winne, 1995). Unfortunately, students are generally not taught these strategies (Peterson & Swing, 1983; Stevens, Slavin, & Farnish, 1991), even though direct instruction of these strategies would likely result in higher achievement (Manion & Alexander, 1997).

Even when instructors teach learning strategies, students may not transfer those strategies to other situations. Research has demonstrated that transfer is most likely to occur when the focus of study has more depth than breadth (Brophy, 1992), which seems problematic in introductory courses that cover a wide array of topics. However, there is hope for the standard Introductory Psychology course. Students can use the concepts from the chapter on learning and memory to enhance their learning of the material. Similarity between the content (what they are learning) and the task (what they need to do) should facilitate transfer (Bassok, 1990). Thus, instructors in this course have an excellent opportunity to develop learning strategies among their students.

Targeting first-year students is important because those who do not arrive at school with sufficient learning skills may be at higher risk for failure (Wood et al., 1999). Which learning strategies are necessary for success is a subject-specific question (Weinstein, Meyer, Husman, Van Mater Stone, & McKeachie, 1999). The best strategies in psychology may not be the best strategies in English or mathematics, and instructors are in a good position to know what strategies are best suited for their classes. However, a good deal of learning strategy research involves adding a study skills course to student schedules (see Petrie & Helmcamp, 1998), adding a learning strategy assignment to class requirements (see Sweidel, 1996), or examining the correlation between variables such as hours of study and grade (see Rau & Durand, 2000). Fleming (2001) demonstrated that making students aware of course-specific tools and strategies at the start of the semester can improve project grades and attitudes. The question remains: Can teachers raise the learning skills of students without adding additional credits to their schedule and without interfering with the curricular agenda?

In this study I investigated whether strategy learning, goal setting, and performance reporting have positive effects on exam performance. I hypothesized that students who have successfully completed at least 1 year of postsecondary education would be less affected by these activities, as I assumed that they had already acquired tools for success. I also expected that students exposed to information on learning strategies as part of the standard curriculum of the course would be able to transfer that knowledge to their learning of the class material.

Method

Participants

Sixty-seven students enrolled in two sections of an Introductory Psychology class participated. Two students subsequently dropped the class, leaving 65 participants. Thirty-one were first-year students, 20 were sophomores, 8 were juniors, and 6 were seniors. Student mean age was 19.77 years (SD = 1.19).

Materials

Demographic information form. This form asked respondents about academic standing, age, sex, status as full- or part-time students, hours of work outside of school, background in psychology, and background in study skills training.

Psychology pretest. This test gauged prior didactic knowledge in the field of psychology with 32 multiple-choice items, selected from the test bank that accompanied the course textbook.

Goal-setting forms. These forms queried students on three learning activities: reviewing notes, reading, and study session. These forms were the same as the goal-setting forms, except the strategies were written in the past tense. To complete the goal-setting form, students indicated the number of minutes they planned to spend on each activity and circled the box containing the strategy they planned to use.

Activity report forms. These forms were the same as the goal-setting forms, except the strategies were written in the past tense. To complete the activity report form, students indicated the actual time spent on each strategy.

Four unit exams. Exams contained 50 multiple-choice items, worth one point each, selected from the test bank.

Design and Procedure

I examined information on demographic and pretest variables to assure that students were randomly divided between the two class sections. Two-tailed t tests (assuming unequal variances, p = .05) revealed the groups were equivalent on the following variables: year in school, age, sex, status (full-time vs. part-time), transfer student (yes or no), hours of employment, background in psychology, prior study skills training, and performance on the pretest. I randomly assigned one section to the treatment condition and one section to the control condition.

During the 16-week semester, each section met for 75 min, 2 days per week. I conducted both sections using the same course packet handouts, PowerPoint® slides, video clips, and music. A blind reviewer observed a random 25% of class periods, examining content, time spent on concepts, de-
livery of materials, and teaching style. He found no differences between the two sections.

During the first day of class, students completed the demographic form and pretest. During the last 5 min of the next six class periods, I gave students in the treatment condition activity report forms and goal-setting forms. I discussed reviewing material strategies on the first day, reading strategies on the second day, and studying material strategies on the third day. I handed out forms with no discussion the next three days. I gave students in the control condition brainteasers on all six days. On the seventh and last day of the first unit, students took the exam.

At the end of the first day of the second unit, I gave students in the treatment condition goal-setting and activity report forms. At the end of each of the next five classes, students turned in the completed form for that day. I handled the daily collection of the materials as a procedural matter at the end of class and did not discuss it. Likewise, I used the last few minutes of class in the control group section to handle the procedural collection and disbursement of assignments and papers.

In both sections, I covered the chapter on learning and memory during the last class of Unit 2, one class period prior to the second exam. The chapter included standard material found in most general psychology textbooks. Unit 3 began 1 week later. I reminded students in both sections of the topics from the previous chapter and encouraged them to use the concepts during the third unit.

Results

Exam Performance

A summary of exam performances appears in Figure 1. Four separate 2 (treatment condition) × 2 (freshmen or upperclassmen) ANOVAs were run to explore exam performance, with a Bonferroni-corrected alpha level of \( p < .0125 \). Means were separated by Fisher's least significant difference method, and differences were considered significant at \( p < .0125 \). Table 1 shows the means and standard deviations for exams.¹

Discussion

The patterns of performance depicted in Figure 1 warrant discussion. First-year students in the treatment condition performed similarly to the upper class students on all four exams, whereas the first-year students in the control condition did not perform as well as their non-first-year classmates or either treatment group on Exams 1 and 2. These exams followed the strategy training intervention, and so the results are not surprising. Entering Unit 3, students in both conditions had learned strategies for successful learning, and I reminded students to use them. In effect, both conditions experienced a treatment of sorts. Not surprisingly, the first-year control students demonstrated significant improvement on Exam 3, and the differences between them and the treatment groups diminished to a nonsignificant level.

What happened next seems to exemplify the concept of transfer difficulty for the first-year control students. They failed to carry over the skills they learned and used in Unit 3 to Unit 4. Their performance went down, whereas the other three groups improved from Exam 3 to Exam 4, creating a statistically significant difference between first-year control students and the other groups on Exam 4 performance. What might explain this finding? The treatment group members may have become comfortable with their new tools for learning. The upper class control group may have had more thoroughly developed scripts for learning already in place, and thus the information on learning presented to them in the class was more readily transferable to Unit 4.

Two key points are supported by this study. First, study strategy instruction can improve exam performance of first-year students. This benefit seems to be limited to first-year students. It makes sense that the performance of first-year students in the treatment condition resembled the upper class students. The treatment group experience seemed to give them a head start in learning skills.

Second, teaching these strategies to students is an investment that seems to pay off beyond the time of treatment. First-year students exposed to the treatment condition outperformed the first-year students in the control condition on the final exam. The final exam took place 13 weeks after the instruction on learning strategies, 8 weeks after they filled out their last goal-setting and reporting forms, and 8 weeks after the curricular memory lesson.

There are limitations to this study, however. Participants were in self-selected, intact groups. The personal data forms indicated no significant differences between the students enrolled in the two sections. However, the classes met at differ-

¹A MANOVA would have been the preferred statistical method; however, there were several students who did not take one of the exams. I opted for a method that enabled me to keep their data and statistically addressed the problem by using the Bonferroni corrected alpha level.
ent times of day: the control group at 11:00 a.m. and the treatment group at 2:00 p.m. Kleitman (1982) hypothesized that the basic rest–activity cycle may be differentially influential on the groups, based on the Time of Day × Attention Demands interaction. This time difference would have been a bigger issue had the control group met in the afternoon, during people’s cycles of a more restful time. The finding that group differences occurred despite the time variable only enhances the seeming validity of the treatment.

Another design issue is that I was the instructor for both sections and was not blind to the group assignment. I took two steps to minimize this limitation. First, a blind reviewer found no differences in my teaching. Second, I delivered presentations using the same materials in both sections.

This study offers implications for intervention. First, students who learn how to be learners maximize their chance for academic success, and although students seem to learn strategies as they move beyond their first year of school, strategy learning can also be hastened by early intervention. This intervention need not create any additional burden on student schedules nor on instructor tasks. Fortunately, a few minutes of class time invested early in the semester can make a substantial difference on first-year student performance throughout. In this case, I taught the strategies in less than 15 min spread over three class periods. In one half of the intervention, I used no in-class time beyond the typical housekeeping period of distributing and collecting papers. Students knew that they had to complete the forms. They also knew I was not going to evaluate the forms. I simply tracked when students completed the forms—I did not grade or process them in any other way.

Second, study skills training can be supplemented, but not replaced, by didactic coverage of memory and learning strategies embedded in the curriculum. Instructors of Introductory Psychology courses might teach learning and memory at the beginning of the term, with the hope that students will personally benefit from understanding the process of memory. This study provides support for such practice. However, results imply that to maximize the potential benefit, teachers should supplement the instruction with didactic coverage of specific learning strategies, as a way to encourage students to set goals and evaluate their success at meeting those goals.

In conclusion, students may not arrive at college prepared for the academic challenges they face. An ineffective period of trial and error may eventually lead students to what they need to do to succeed. Instructors of first-year students are in the important position of influencing this transition for students. It is encouraging that such minimal effort on the part of instructors can have an impact on student success.

References


### Table 1. Exam Scores for First-Year and Upper Class Students With and Without Training

<table>
<thead>
<tr>
<th>Exam</th>
<th>Treatment</th>
<th></th>
<th></th>
<th>Control</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td>39.70</td>
<td>4.96</td>
<td>40.19</td>
<td>3.97</td>
<td>35.82</td>
<td>4.96</td>
</tr>
<tr>
<td>2</td>
<td>39.45</td>
<td>3.87</td>
<td>41.25</td>
<td>3.86</td>
<td>34.45</td>
<td>4.96</td>
</tr>
<tr>
<td>3</td>
<td>39.10</td>
<td>5.23</td>
<td>39.67</td>
<td>2.61</td>
<td>35.82</td>
<td>4.96</td>
</tr>
<tr>
<td>4</td>
<td>44.40</td>
<td>3.90</td>
<td>43.23</td>
<td>5.60</td>
<td>39.20</td>
<td>7.02</td>
</tr>
</tbody>
</table>

*Different from non-first-year control scores, significant at p < .0125, and different from both groups in the treatment condition, significant at p < .05. Different from non-first-year treatment scores, significant at p < .05. Different from first-year treatment scores, and non-first-year control scores, significant at p < .05. Different from first-year treatment scores, and non-first-year control scores, significant at p < .0125.
Personal Ad Content Analysis Teaches Statistical Applications

D. W. Rajecki
Indiana University–Purdue University Indianapolis

Students analyzed newspaper personal advertisements for the age preferences of younger and older male and female writers. Several applications of the \( F \) statistic revealed between-group preferences for partners of relatively different ages. Many applications of the \( r \) statistic revealed within-group preferences for partners of absolutely similar ages. Students communicated this formal statistical information in a written report. Subsequently, most students rated the personal ads as interesting and as giving the impression of working with real people. Success on final examination computational items (including naming degrees of freedom and probabilities) exceeded 90%. A separate quantitative assessment revealed a reliable increase in mastery of \( r \) and \( F \) statistical notation and concepts from before to after the project.

Personal advertisements yield convenient samples of inherently interesting societal raw data. Researchers gather such material from national tabloids (Harrison & Saeed, 1977), local newspapers (Rajecki, Bledsoe, & Rasmussen, 1991), and the Internet (Matthews, 1999). Ad content analysis provides beginning students with useful exposure to basic statistical applications and procedures. Identifying ads as an engaging pedagogical tool contributes to the growing literature on methods and issues in statistical instruction (cf. Friedrich, Buday, & Kerr, 2000; Ware & Brewer, 1999; Ware & Johnson, 2000).

My technique is based on a recent study by Rasmussen et al. (1998) that examined the partner age preferences of ad writers whose own ages ranged from the 20s to 50s. Those researchers reported as many as 30 separate correlation coefficients (\( r \)) indicating that across the life span male and female advertisers consistently sought others of roughly similar age. At the same time, \( F \) tests indicated that women generally preferred somewhat older partners, whereas men preferred somewhat younger partners. Finally, in later segments of the life span, writers of both sexes tended to prefer progressively younger partners.

For my purposes, I scaled down the Rasmussen et al. (1998) design to give undergraduates meaningful and manageable opportunities to work extensively with the \( r \) and \( F \) statistics. I created opportunities for mechanical practice with computer data entry and computation. Writing assignments (including tables) provided experience with the proper communication of statistical information such as degrees of freedom and probability estimates.

In the classroom, this project can certainly be couched in theoretical terms. For example, are people's mate preferences better explained by concepts from evolutionary theory (Kenrick & Keefe, 1992) or social exchange theory (Rasmussen et al., 1998)? Alternatively, students can empirically address a number of interesting sets of questions. For example:

1. Compared with one another, do men and women in different age decades (say, the 20s, 30s, or 40s) have different partner age preferences?

2. All else being equal, do men and women prefer partners with ages similar to their own ages?

The first set of questions can be answered by evaluating group mean differences with the \( F \) statistic. The second set of questions can be answered by showing the strength of association between own and partner ages with the \( r \) statistic. In my classes, I gave students practice with both types of tests: They calculated two separate ANOVAs and 16 separate correlation coefficients.

Notes

1. I thank Josh Severson for his help in coding the reliability of teaching methods and Andrew P. Manion for consultation on an earlier draft of this article.

2. Send correspondence and requests for sample goal setting and activity forms to Victoria Manion Fleming, 201 McGuffey Hall, Miami University, Oxford, OH 45056; e-mail: V.Fleming@muohio.edu.
We examined how academic background and course involvement differentially predicted students' performance on lecture- and text-based exam questions (N = 114; 34% men; 76% freshmen). Results showed that academic background and course involvement predicted performance on lecture-based questions and overall exam performance, whereas academic background variables alone predicted performance on text-based questions. We discuss the pedagogical implications of the findings on course and test design.

Although research has documented a link among student background, course involvement, and exam performance, few studies have examined their association with exam performance in text-based courses versus more lecture-based courses (where students’ performance is based primarily on lecture materials not found in the textbook) or courses that place equal emphasis on lecture and text information. One study (Hardy, Zamboanga, Thompson, & Reay, 2003) examined academic background variables and course involvement and their association with exam performance in an Introduction to Psychology course that was primarily text-based. Findings showed that academic background variables such as academic aptitude (as measured by the American College Test [ACT]), prior grades, and past psychology courses, but not course involvement such as lecture attendance and performance in the discussion section of the course, were positively associated with exam performance. The researchers noted that the text-based design of the course may have explained why exam performance was unrelated to lecture attendance and discussion section performance. Unfortunately, Hardy et al. (2003) did not separate exam performance by text-based versus lecture-based questions. This study limitation raises the possibility that students who attended class regularly in a text-based course might have performed well on exam questions derived from lecture materials not covered in the text (e.g., guest lectures, video presentations, in-class demonstrations). Additionally, the authors measured
lecture attendance via self-reports, raising the question of possible inaccuracies in their assessment of course involvement.

It is common practice for instructors to encourage their students to attend lectures as well as to participate in the discussion sections of a given course to derive the maximum benefit for their learning. However, this question remains: Does student involvement in these activities facilitate course performance? If so, what kind of student is likely to benefit from this practice, and do course and test design matter? To address these research questions, we examined which aspects of academic background and course involvement were associated with performance on lecture- and text-based exam questions in an Introduction to Psychology course. This investigation builds on prior research and addresses limitations found in previous studies (Beecher & Fischer, 1999; Hardy et al., 2003; Thompson & Zamboanga, 2004). Although prior studies have highlighted the utility of academic background variables in predicting course performance, they have not addressed how academic background and course involvement variables might be differentially associated with student performance on lecture- versus text-based exam questions. Moreover, with respect to the link between lecture attendance and performance on lecture- versus text-based quiz questions, researchers (e.g., Shimoff & Catania, 2001) did not account for the influence of other known correlates (e.g., prior aptitude or knowledge) of academic achievement. To address these limitations, we separated exam questions into lecture- and text-based questions. Moreover, to account for possible overlap between text and lecture materials on the exam, we differentiated text- and lecture-based questions that were mutually exclusive of each other. That is, we identified lecture-based questions that students could not answer based on material in the text and vice-versa. Doing so allowed us to better investigate the influence of course involvement on exam performance and the relevance of test design in evaluating student performance, and to infer the relevance of course design in student achievement. Consistent with the Hardy et al. (2003) study, we assessed course involvement via lecture attendance and discussion section performance (as indexed by discussion section attendance, exercise activities, and a quiz). Additionally, instead of relying on self-reported lecture attendance, we used instructor records to assess this variable. Because academic background variables are associated with more enduring contributions to course performance than course involvement (Jansen & Bruinsma, 2005), we expected background variables to be predictive of performance on both lecture- and text-based exam questions. However, we anticipated that course involvement (class attendance) would be predictive of exam performance on lecture-based questions, whereas academic background variables would be predictive of exam performance on text-based questions.

Method

Sample

Participants were students enrolled in an Introduction to Psychology course (taught by one of the investigators of this study) at a Midwestern state university (N = 114; 34% men; 76% freshmen, 14% sophomores, 5% juniors, 5% seniors). Students met biweekly for lectures and weekly in a small-enrollment discussion section. The lecture section entailed presentation of new information, whereas the discussion section (two sections, each taught by a graduate teaching assistant) focused on in-depth discussions of relevant course materials and lecture topics in a smaller group setting. We used instructor records to assess year in school, gender, attendance, discussion section performance, and exam scores for this study. Prior to taking the final exam, students completed a self-report questionnaire in which they reported on their academic background. The institutional review board at the university where we conducted this study approved these procedures.

Measures

Student background. These measures included self-reported ACT score, grade-point average (GPA), number of prior psychology courses taken, gender, and school year.

Exam performance. We computed a total exam performance score by averaging students’ scores on four exams (α = .88). The instructor also carefully reviewed each exam question and identified which test items were based exclusively on materials covered in lectures or the text. The instructor combined and averaged all exclusively lecture-based exam items (α = .77; n = 40, 20% of all test items) and applied the same procedure to all exclusively text-based questions (α = .73; n = 44, 22% of all test items). The remaining exam items consisted of test questions derived from both lecture and text.
Discussion section performance. We derived a discussion section performance score by summing the number of discussion sections attended and the homework and test scores for each student. We standardized these variables prior to creating our discussion section performance score.

Lecture attendance. We derived lecture attendance by summing 13 random periodic assessments conducted by the instructor throughout the course of the semester.

Results

Descriptive statistics and correlations appear in Table 1. Being female was associated with higher class year and higher prior GPA. There were positive associations between ACT and total exam performance, performance on text-based questions, and performance on lecture-based questions. Prior GPA was positively associated with ACT score, lecture attendance, total exam performance, text-based question performance, and lecture-based question performance. Discussion section performance and lecture attendance were positively associated. Lecture attendance was positively associated with total exam performance, performance on text-based questions, and performance on lecture-based questions.

We tested the study hypothesis using the following three hierarchical linear regression analyses: one model to assess predictors of overall exam performance, a second model to predict performance on exclusively lecture-based questions, and a third model to predict performance on exclusively text-based questions (see Table 2). For each regression analysis, we entered gender and class year as demographic controls and entered the key study variables (ACT score, prior GPA, past psychology courses, discussion section performance, and lecture attendance) as predictors. The first regression model tested the relative role of student background and course involvement variables in predicting overall exam performance, \( R^2 = .45, p < .01 \). ACT score, prior GPA, and lecture attendance were all positively associated with overall exam performance. However, the number of prior psychology courses taken and discussion section performance were unrelated to overall exam performance. We assessed the comparative role of student background and course involvement variables in predicting performance on lecture-based questions in the second regression model, \( R^2 = .39, p < .01 \). ACT score, prior GPA, and lecture attendance were all positively associated with performance on lecture-based questions. However, the number of prior psychology courses taken and discussion section performance were also unrelated to performance on lecture-based questions. The third regression model assessed the relative role of student background and course involvement variables in predicting performance on text-based questions in the second regression model, \( R^2 = .33, p < .01 \). ACT scores were positively related to exam performance on text-based questions. However, prior GPA, the number of prior psychology courses taken, discussion section performance, and lecture attendance were all unrelated to performance on text-based questions.

The sample for this study (\( N = 114 \)) came from a larger sample of 193 students from a psychology course. However, 79 of these students had missing data on at least one of the study variables; therefore, we excluded...
Table 2. Hierarchical Multiple Regression Models Predicting Total Exam Performance, Lecture-Based Questions, and Text-Based Questions

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Total Exam Performance</th>
<th>Lecture-Based Questions</th>
<th>Text-Based Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>Beta</td>
</tr>
<tr>
<td>Gender*</td>
<td>−.35</td>
<td>.86</td>
<td>−.03</td>
</tr>
<tr>
<td>Yearb</td>
<td>.68</td>
<td>.50</td>
<td>.10</td>
</tr>
<tr>
<td>ACT</td>
<td>.57</td>
<td>.11</td>
<td>.40**</td>
</tr>
<tr>
<td>Prior GPA</td>
<td>1.76</td>
<td>.74</td>
<td>.21**</td>
</tr>
<tr>
<td>Prior Psychology Courses</td>
<td>.20</td>
<td>.58</td>
<td>.03</td>
</tr>
<tr>
<td>Discussion Section</td>
<td>−.02</td>
<td>.50</td>
<td>−.01</td>
</tr>
<tr>
<td>Lecture Attendance</td>
<td>.73</td>
<td>.20</td>
<td>.31**</td>
</tr>
</tbody>
</table>

Note. N = 114. \( R^2 = .45 \) (p < .01) for Total Exam Performance; \( R^2 = .39 \) (p < .01) for Lecture-Based Questions; \( R^2 = .33 \) (p < .01) for Text-Based Questions. All betas are standardized and based on the final model. *Gender was coded women = 0, men = 1; bYear was coded freshman = 1, sophomore = 2, junior = 3, senior = 4. *p < .05; **p < .01.

them from the analyses. ANOVAs of the study variables comparing students included in the sample to those excluded due to missing data showed that those with missing data reported significantly lower levels of discussion section performance, lecture attendance, and exam performance. Thus, to assess the impact of missing data on the primary analyses, we estimated the regression model parameters a second time, this time using multiple imputation procedures in SAS 9.1 to account for missing data. The pattern of results did not differ from that reported for the smaller sample with complete data; thus, we maintained the results for the sample with complete data.

Discussion and Teaching Implications

We examined which aspects of academic background and course involvement are associated with performance on lecture- and text-based exam question. First, results showed that academic background (ACT) and course involvement (lecture attendance) were predictive of overall exam performance and performance on lecture-based exam questions in an Introduction to Psychology course. Prior aptitude (as measured by ACT) was the only significant predictor of performance on text-based questions. These findings indicate that achievement (prior GPA) may be less important than aptitude (ACT) in predicting performance on text-based questions. Given our findings, student performance on the ACT continues to be a relevant predictor of college academic performance. Second, our results are inconsistent with the Shimoff and Catania (2001) study in which lecture attendance was linked to performance on text-based exam questions. In our study, we accounted for any overlap between text and lecture materials on the exam by incorporating only text- and lecture-based questions that were mutually exclusive of each other. It remains unclear whether Shimoff and Catania accounted for such factors, which may help explain the discrepancies in our results.

Our findings, coupled with the results obtained by Hardy et al. (2003), suggest that the association between academic background variables, course involvement, and course achievement may differ as a function of how instructors design their course and assess exam performance. Such findings have implications for the different pedagogic methods used by instructors, as well as the different learning styles, student background (e.g., nontraditional students), and specific types of classes, including distance learning courses where students receive a textbook but may have limited access to or contact with the instructor. For example, instructors should consider how they might evaluate student performance and recognize the possibility that highly text-based courses may be more conducive to course achievement for students with high levels of scholastic ability regardless of their class attendance. Conversely, in a class that is not primarily text-based, course involvement such as lecture attendance is an important predictor of exam performance, particularly in courses in which instructors evaluate students on their knowledge of lecture material. However, class attendance may not be as beneficial to students enrolled in classes where lecture material overlaps substantially with material in the text and instructors do not evaluate student learning beyond the book content. In these cases, lectures may add little to student learning and class attendance is less relevant to student achievement; thus, we encourage instructors to teach additional
information in class to make attendance more conducive to student learning.

Contrary to our expectations, we did not find any significant associations between discussion section performance and exam performance. This finding is consistent with that of Hardy et al. (2003) and suggests that discussion section performance is unrelated to exam performance in both text- and lecture-based courses. One possible explanation for this finding is that the material presented in the discussion sections did not always overlap with information in the text or lecture. It remains unclear whether or how the current findings would have differed had the discussion sections focused primarily on materials found in the text and information covered in the lecture. In light of these findings, discussion sections may be of questionable value unless they provide activities that directly enhance student performance, particularly as assessed on exams or other outcome measures. Thus, in designing classes, we encourage instructors to carefully consider the purpose and utility of discussion sections as they relate to student learning and achievement.

There are some limitations and future research directions worth noting regarding our study. First, the cross-sectional design of this investigation precludes any inferences of causality regarding the associations between the study variables. As such, experimental studies comparing the role of student background and course involvement across a variety of course designs and performance evaluation techniques are needed. Furthermore, this study examined a relatively homogeneous sample of traditional college students. Given the rapid growth in numbers of nontraditional students on college campuses (Chao & Good, 2004), it would be interesting for future research to assess predictors of course performance among a more diverse sample (including nontraditional students) in an attempt to draw conclusions that might be important for different student populations. Interestingly, we measured attendance quite differently from Hardy et al. (2003)—random periodic assessments conducted by the instructor versus self-report—yet our findings paralleled the results obtained in their study. Perhaps self-reports of attendance are more accurate than researchers might assume and, as such, future studies could closely test this observation. Finally, when examining issues of course design, it would be interesting to obtain student perceptions on the type of course they believe is most beneficial to their learning (text- vs. lecture-based courses) and its relevance to academic background and student achievement.

References


Notes

1. We are grateful to Monica Barnard, Jennifer A. Haase, Sarah Kepple, Sheryl Mitchell, Chris Nielsen, Chris Noffke, Grant Orley, Eric Owens, Sherill Pineda, and Traci Swinnea for their assistance with this project. Special thanks to Janine V. Olthuis for her editorial assistance on the manuscript. Sam A. Hardy is now at Brigham Young University.

2. Send correspondence to Byron L. Zamboanga, Clark Science Center, 44 College Lane, Department of Psychology, Smith College, Northampton, MA 01063; e-mail: bzimmerman@smith.edu.
Effects of Test Expectation on Multiple-Choice Performance and Subjective Ratings

William R. Balch
*The Pennsylvania State University, Altoona*

Undergraduates studied the definitions of 16 psychology terms, expecting either a multiple-choice (n = 132) or short-answer (n = 122) test. All students then received the same multiple-choice test, requiring them to recognize the definitions as well as novel examples of the terms. Compared to students expecting a multiple-choice test, those expecting a short-answer test performed similarly on example questions but significantly better on definition questions. Students in these 2 test-expectation conditions also differed in several subjective ratings of their study and test taking. The results suggest that students do not typically study in an optimal way for multiple-choice tests.

Because instructors can score multiple-choice tests quickly, conveniently, and objectively, these tests are common and often necessary in teaching large sections of courses such as introductory psychology. Multiple-choice tests require only recognition, not recall, of correct information. Thus they are popular with students, who perform better on multiple-choice tests than on recall tests of the same material (e.g., Pajares & Miller, 1997). Students also rate the similarity between multiple-choice and recall tests as low (Rocklin, 1992).

Despite the frequent use of multiple-choice questions, there is little recent research on how students actually study for such tests. The most pertinent objective evidence comes mainly from a few experiments performed more than 20 years ago. One such study (d’Ydewalle, Swerts, & de Corte, 1983) suggested that, compared to participants expecting recall tests (e.g., short-answer, fill-in-the-blank, essay tests), participants expecting multiple-choice tests did not perform as well on either type of test. Participants studying textbook excerpts performed 10% to 15% better on a recall test when expecting that kind of test than when expecting a multiple-choice test. Participants expecting the recall test also performed 5% better on a multiple-choice test: a smaller but significant difference. In addition, they reported a longer study time when expecting a recall test as compared to a multiple-choice test. These results suggest that normally students do not prepare for multiple-choice tests in such a way as to maximize their test performance.

However, not all researchers have found that a multiple-choice expectation leads to poorer test scores. In one study, students expecting multiple-choice questions performed better on a multiple-choice final exam than those who expected a recall final exam (Sax & Collet, 1968). In other studies, participants performed equally well regardless of the type of test they expected (Kumar, Rabinsky, & Pandey, 1979; Vallance, 1947).

One reason for the variable outcomes of research on students’ test expectations may be that the experiments have involved different student populations taking tests on material from different disciplines. For instance, d’Ydewalle et al. (1983) tested first-year law students on material relating to the history of the Roman Empire; Sax and Collet (1968) tested undergraduates in a tests-and-measurements class who were studying for a final exam; and Kumar et al. (1979) tested ninth graders on geographic, historical, and economic aspects of the Republic of Sudan. Moreover, students in those experiments generally paced their own study and studied passages of connected discourse rather than discrete items. Thus, these experiments varied in their control over the presentation of the material to be tested and in the content as well as the difficulty level of the test questions. To address the aforementioned problems, I had undergraduates in both a multiple-choice
and a short-answer test-expectation condition of my experiment study the definitions of psychology terms one at a time and at a uniform, experimenter-paced rate. I also paced students through a multiple-choice test that required them to recognize both the definition and a novel example of each term. In this way, I could assess the effect of test expectation on two levels of multiple-choice test performance. The definition questions required only rote memory (Bloom’s first-level educational objective: “knowledge”), and the example questions required comprehension of the definitions (Bloom’s second-level objective: “comprehension”; Bloom, 1956).

I hypothesized that, compared to students expecting a multiple-choice test, students expecting a short-answer test would perform better on definition questions (Hypothesis 1) as well as on example questions (Hypothesis 2) of a multiple-choice test. In other words, I believed that a short-answer test expectation would help students study with more retention and also with more comprehension. The rationale for my study was to assess, under reasonably controlled yet still partially realistic experimental conditions, how students might normally study for multiple-choice tests in introductory psychology courses. This type of knowledge could suggest ways for students to prepare more effectively for such tests. Finally, as additional but subjective measures of the effects of test expectation, I had students rate several aspects of their study and test taking.

Method

Participants

Two hundred fifty-four undergraduates (145 women and 109 men) in two different sections of introductory psychology voluntarily participated (for extra credit) in this experiment during the Fall 2006 semester. The experiment occurred before the presentation (in lectures or textbook assignments) of any of the psychology terms included in the experiment.

Materials

Study booklets. In every study booklet, the definitions of 16 psychology terms appeared twice in consecutive random orders. Each definition appeared on a separate page, and the resulting sequence of 32 pages was the same for all booklets. Four of the terms represented each of four different psychology topics: experimental method (independent variable, dependent variable, control condition, confounding variable), parts of the brain (medulla, hypothalamus, thalamus, reticular formation), psychological disorders (somatoform, dissociative, schizophrenic, bipolar), and psychotherapies (client-centered therapy, systematic desensitization, psychoanalysis, rational-emotive therapy). I adapted these definitions from those listed in the introductory psychology textbook that I use (Bernstein, Penner, Clarke-Stewart, & Roy, 2006). However, all definitions were about the same length, did not refer to other related terms, and took the form of complete sentences (e.g., “A somatoform disorder is a psychological condition in which a person has the symptoms of a physical problem, without a physical cause”; Bernstein et al., 2006, p. 595).

On each page of every study booklet, students indicated a subjective comprehension rating on a scale from 0 (lowest) to 10 (highest) that appeared below the definition printed on that page. Each of the 32 ratings (2 for each of the 16 definitions) reflected perceived understanding of a given term, based on the information provided on that page. I included this task to provide a subjective measure of comprehension and to help students maintain their attention as they studied the booklets. Because students in both test-expectation conditions performed the same rating task, the ratings should not have differentially influenced the studying in these two conditions.

Test booklets. All test booklets consisted of the same multiple-choice questions. One question appeared on each page and the questions occurred in one of two different random orders (Form A or Form B). Sixteen were definition questions, beginning with “The definition of [a given term] is . . .” followed by four choices. The correct choice was the definition of the term as it had appeared in the study booklets, whereas the incorrect choices were the definitions of the three related terms. On the definition question for somatoform disorder, for instance, the definition of that term—along with the definitions of the three other psychological disorders (dissociative, schizophrenic, bipolar)—were the four choices.

The other 16 questions were example questions that began with “An example of [a given term] would be . . .” On the example question for somatoform disorder, for instance, the correct example was “a person who frequently sees doctors, complaining of illnesses that there is no medical evidence of.” The three incorrect choices were examples of the other three psychological disorders (dissociative, schizophrenic, bipolar).
In addition to the 32 multiple-choice questions, the test booklets also included six subjective rating scales that students completed after finishing the multiple-choice test. These scales, in the order presented in the booklet, were perceived test performance, from 0 (very poor) to 10 (very good); study time, from 0 (only as long as it took to read each definition in the study booklet) to 10 (the total allotted time); attention to repeated definitions (i.e., the second time each given definition appeared in the study booklet), from 0 (no much attention) to 10 (a lot of attention); amount of elaboration during study, from 0 (trying to remember mainly the words of each definition) to 10 (trying to think of extra ways to remember each definition); anxiety during study, from 0 (no anxiety) to 10 (very high anxiety); and anxiety during test, from 0 (no anxiety) to 10 (very high anxiety).

Procedure

In each of two morning sections of introductory psychology that I taught, I randomly assigned each student to one of two test-expectation conditions: short-answer or multiple-choice. I performed the experiment during the first week of classes, before any regular coursework began. During the first class of the semester, I displayed in each section a list of the names of students in each condition (arbitrarily called the Blue or the Green group). Then I conducted an experimental session with just one of the test-expectation groups during the remainder of the first class. During the second class that week, I conducted a session with the other group. I counterbalanced across the two sections the assignment of the test-expectation groups to the first or second class. Students in the multiple-choice test-expectation condition received instructions that they would take a multiple-choice test. Their score, I said, would reflect how well they understood the meaning of the terms that they had studied. However, I instructed students in the short-answer test-expectation condition that they would take a written short-answer test on the terms. I told these students that they would need to explain the terms that they had studied in answers of one or two sentences each and that they would be scored on how well they had understood the meaning of each term.

In each test-expectation condition, I instructed students that they would be studying a booklet presenting the definitions of 16 psychology terms—one at a time—and that the definition of each term would appear twice somewhere in the booklet. I also told all students that, on each page, there would be a scale from 0 (lowest) to 10 (highest) on which they would be rating their comprehension of the definition on that page. Finally, I instructed students in both conditions that they should study the terms conscientiously and do as well as they could on the test. However, I assured them that I would use the data from this experiment for research purposes only and that their test scores would in no way influence their introductory psychology grades.

After these initial instructions, students studied each numbered page of the study booklet until I said, “Rate,” at which time they rated their comprehension for that page. When I said, “Turn to [whatever page number was next],” they turned the page, studied the definition on that page, and so forth. In pacing students through their booklets, I allotted 30 sec for each page: 20 sec for study and 10 sec for completing the comprehension rating on that page.

As soon as students finished the study booklets, I collected these booklets and distributed the aforementioned Form A or Form B test booklets by random assignment. In both test-expectation conditions, I instructed students that they would be taking a multiple-choice test as part of the design of the experiment and that I would explain the experiment more fully after they finished the test. I also told them that on some of the multiple-choice questions, they were to recognize the definition of a previously studied term. On others, they were to indicate an appropriate example of the term. On each page of the test booklet, they were to read the question on that page, answer it without waiting for a separate instruction, and then wait for instructions to turn the page. The interval between students’ finishing their study booklets and their starting the test was about 5 min. I paced them through each page of the test booklet, allowing 35 sec per page. After they finished all the multiple-choice questions, I instructed them to complete the aforementioned subjective rating scales. I paced students through the ratings one scale at a time, explaining each rating briefly and allowing about 12 sec for their completing each rating.

Results

For each of the dependent measures—objective (definition and example) scores as well as the subjective ratings—I performed separate independent-samples one-way ANOVAs of the difference between means of the short-answer and multiple-choice test-expectation conditions. Table 1 shows the means and
Example questions. The scores on the example questions were generally lower than scores on the definition questions, but there was no significant difference between students in the short-answer (61.4%) and the multiple-choice (59.5%) test-expectation conditions, $p > .10$ (see Table 1). In other words, the short-answer test expectation did not improve performance on the relatively challenging example questions. These questions required students to extrapolate beyond the information in the study booklets.

Subjective Ratings

Ratings completed during study. Table 1 shows the average comprehension rating for each test-expectation condition, based on a scale from 0 (lowest) to 10 (highest). In contrast to the pattern of results with the objective scores, students expecting the multiple-choice test averaged significantly higher (8.18) in their comprehension ratings than did students expecting the short-answer test (7.39), $F(1, 252) = 26.00, p < .001, \eta^2 = .09$. This result occurred even though the former students did not perform significantly better on example questions than did the latter students and performed significantly worse on definition questions.

Ratings completed after the test. For three of these ratings, there were significant differences between students in the two conditions (see Table 1). First, students expecting the short-answer test rated their attention to repeated definitions (i.e., the second time that each definition appeared in the study booklet) higher (5.85) than did students expecting the multiple-choice test (5.12), $F(1, 252) = 8.46, p < .01, \eta^2 = .03$. Next, students in the short-answer test-expectation condition averaged higher (4.65) on their anxiety during study ratings than did students expecting a multiple-choice test (3.68), $F(1, 252) = 10.31, p < .01, \eta^2 = .04$. However, the former students averaged lower (3.07) on their anxiety during test ratings than did the latter students (4.11), $F(1, 252) = 12.33, p < .001, \eta^2 = .05$.

For the other three ratings, the differences between students in the two test-expectation conditions were not significant. The amount of elaboration during study ratings were similar for students expecting a short-answer (5.54) or a multiple-choice (5.33) test, $F < 1, p > .10$. Students expecting a short-answer test rated their study time (6.56) and their perceived test performance (5.93) slightly higher than did students expecting a multiple-choice test (6.06 and 5.49, respectively).

### Table 1. Objective Scores and Subjective Ratings as a Function of Test Expectation

<table>
<thead>
<tr>
<th>Measure</th>
<th>Test Expectationa</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short Answerb</td>
<td>Multiple Choicec</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td></td>
</tr>
<tr>
<td>Objective scoresd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definition</td>
<td>77.00*</td>
<td>16.71</td>
<td>71.02</td>
<td>20.76</td>
<td></td>
</tr>
<tr>
<td>Example</td>
<td>61.38</td>
<td>16.25</td>
<td>59.52</td>
<td>16.47</td>
<td></td>
</tr>
<tr>
<td>Subjective ratingsg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensionf</td>
<td>7.39***</td>
<td>1.43</td>
<td>8.18</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>Perceived test performance</td>
<td>5.93</td>
<td>1.87</td>
<td>5.49</td>
<td>1.89</td>
<td></td>
</tr>
<tr>
<td>Study time</td>
<td>6.56</td>
<td>1.99</td>
<td>6.06</td>
<td>2.07</td>
<td></td>
</tr>
<tr>
<td>Attention to repeated definitions</td>
<td>5.85**</td>
<td>1.93</td>
<td>5.12</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>Amount of elaboration during study</td>
<td>5.54</td>
<td>2.39</td>
<td>5.33</td>
<td>2.38</td>
<td></td>
</tr>
<tr>
<td>Anxiety during study</td>
<td>4.65**</td>
<td>2.40</td>
<td>3.68</td>
<td>2.14</td>
<td></td>
</tr>
<tr>
<td>Anxiety during test</td>
<td>3.07***</td>
<td>2.11</td>
<td>4.11</td>
<td>2.31</td>
<td></td>
</tr>
</tbody>
</table>

Note. $N = 254.$

*Based on experimenter’s instructions before students studied the definitions. $b_{n} = 122.$ $c_{n} = 132.$ $d$Percent of 16 multiple-choice questions answered correctly. $e$From 0 (lowest) to 10 (highest). Except as noted, students completed each rating after taking the multiple-choice test. $f$Average of 32 ratings, each made directly after studying the definition appearing on each page of the study booklet.

*p < .05. **p < .01. ***p < .001. Based on independent-samples one-way ANOVAs of the difference between the short-answer and the multiple-choice test-expectation conditions.

standard deviations in each condition. I report as significant those effects for which $p < .05$. For each significant effect, I also report effect size ($\eta^2$).

Objective Scores

Definition questions. As Table 1 illustrates, students expecting a short-answer test scored significantly higher (77%) on the definition questions of the multiple-choice test than did those expecting a multiple-choice test (71%), $F(1, 252) = 6.33, p < .02, \eta^2 = .025$. Thus, compared to the multiple-choice expectation, the short-answer expectation improved performance on the relatively easy definition questions that required students only to remember the same information that appeared in the study booklets. This effect is consistent with the results of d’Ydewalle et al. (1983), who found a comparable test-expectation effect on multiple-choice performance.
but for both ratings the difference was only nearly significant, $F_s \leq 3.42, p < .10$. Note that d’Ydewalle et al. (1983) found that students expecting a recall test reported significantly longer study times than those expecting a multiple-choice test. Possibly because my experimenter-paced procedure might have produced a ceiling effect on study time, I did not find such a result.

Discussion

In support of Hypothesis 1, students expecting a short-answer test performed better on definition questions than did students expecting a multiple-choice test. This finding suggests that students might not normally use the most effective strategies in studying for multiple-choice tests. Although students generally perform better on multiple-choice than on other types of tests (e.g., Pajares & Miller, 1997), they might still underestimate the difficulty of multiple-choice tests.

In addition, the results obtained with several subjective ratings seemed consistent with Hypothesis 1. First, students expecting a short-answer test rated their attention to repeated definitions higher than did students expecting a multiple-choice test. These ratings suggest that, by paying greater attention the second time each term appeared in the study booklets, students expecting the short-answer test might have distributed their study more. Dempster (1988), for instance, discussed the educational benefits of distributed study. Second, the two anxiety ratings provided data consistent with the better definition-question performance of students in the short-answer test-expectation condition. Although none of the means of those ratings exceeded the midpoint of the scales from 0 (lowest) to 10 (highest), the students expecting the short-answer test rated their anxiety during study as higher—but their anxiety while taking the test as lower—than did those expecting the multiple-choice test. Their moderately higher anxiety during study might have motivated students expecting the short-answer test to work harder at learning the definitions. Likewise, their lower anxiety during the test might have reflected greater confidence in their preparation for the test.

I found no evidence for Hypothesis 2, stating that students expecting a short-answer test would perform better on example questions than would students expecting a multiple-choice test. Compared to the definition-question scores, the example-question scores of students in both test-expectation conditions tended to be lower, and there was no significant difference between the conditions. Despite my instructing students expecting the short-answer test that they would need to understand and explain the terms defined in the study booklets, these students performed no better on example questions than did students expecting a multiple-choice test.

The subjective-rating data are consistent with the notion that studying to recall material does not in itself improve comprehension. Students expecting a short-answer test did not report using significantly more elaboration than did those expecting a multiple-choice test. While studying the definitions, in fact, the latter students rated their comprehension significantly higher than did students expecting a short-answer test. Because the students expecting a multiple-choice test performed no better on the example questions and worse on the definition questions, it is unlikely that these students really understood the definitions better than did those expecting a short-answer test. Instead, students expecting the multiple-choice test might have had a lower criterion for the level of comprehension required for answering multiple-choice questions and rated their comprehension accordingly. For instance, a student in the short-answer test-expectation condition commented after the experiment that he would have rated his comprehension higher had he known that he would actually receive "only a multiple-choice test."

My failure to find a test-expectation effect on example-question performance is consistent with the results of a recent experiment suggesting that students’ preparation for even short-answer tests might be limited. Winne and Jamieson-Noel (2003) investigated the study tactics of undergraduates who expected a short-answer test on a meteorological text excerpt. In terms of seven study tactics (e.g., learning terms in boldface, memorizing figure labels, comparing figures to the text), students reported using these tactics less frequently and in different proportions compared to the optimal use that a theoretical model predicted.

Teachers generally want their students to develop thinking and reasoning skills (e.g., Halonen, 1995; Wade, 1995). Nevertheless, most teachers emphasize content in their courses (Benjamin, 2005, p. 147) and often test students with questions that require only memory of facts (McKenzie, 1973, p. 287), such as the definition questions in my experiment. My results suggest that, to the extent that an instructor tests memory for content, his or her students might benefit from
supplementing their study for multiple-choice tests with some recall-based study strategies.

Yet study that involves elaboration, rather than merely an attempt to recall, would promote performance on more difficult test questions. For instance, compared to a nonelaborating control group, students who performed exercises in which they elaborated on textbook material (by giving examples of concepts, answering thought questions, etc.) scored higher on test questions requiring comprehension, reasoning, and applying the concepts to daily life (Watson, Hagihara, & Tenney, 1999). Similar results occur when the instructor provides examples of psychological terms (Balch, 2005) or engages students in elaborating such terms with mnemonics (Carney & Levin, 1998). Finally, prelecture quizzes on textbook assignments improve students’ performance on higher level multiple-choice exam questions and might also improve their preparation for, as well as their comprehension of, lectures (Narlock, Garbin, & Turnage, 2006).

For the purposes of experimental manipulation and control, I used several procedures in my study that might limit its applicability to the way students prepare for real multiple-choice tests. First, students typically do not follow a preset pace, as the students in my experiment did. In addition, I tested students only about 5 min after they studied the definitions of the psychology terms. Except for those engaging in the most extreme form of cramming, students generally take exams testing their retention over a somewhat longer interval. Moreover, the test that my students took did not count toward their grades. In particular, my short-answer test-expectation condition was unrealistic. A teacher would not normally mislead students about the kind of test that they would be taking. Thus, if an instructor wanted to help students by having them try to recall material that they are studying, he or she would need to try other methods. During an office visit from a student, for instance, a teacher could illustrate and recommend a supplemental study strategy in which someone asks a series of questions and the student gives short answers. Although Balch (2001) found no significant correlation between students’ rated use of such a question-and-answer study technique and their course grades, students in the experiment that I report here did perform better on definition questions when preparing as for a short-answer test. A promising direction for future research is the further development of effective and realistic methods that help students prepare better for multiple-choice tests in a way that is relatively labor-efficient for instructors and encourages students to take responsibility for their learning.

References


Notes

1. I thank Ashley M. Smyder for her assistance in conducting, and in scoring the raw data from, the experimental sessions in this study. In addition, I am grateful to George Greiner and Peter Gionet for insights that were helpful in my preparing this article.

2. Send correspondence to William R. Balch, Department of Psychology, The Pennsylvania State University, 3000 Ivyside Park, Altoona, PA 16601; e-mail: wrb3@psu.edu.
These findings have direct implications for instructors and researchers. Instructors may be concerned with unrealistically high grade expectations because “one way to ensure dissatisfaction is to have expectations that one can not easily meet” (Gautney & Cann, 2001, p. 86). One might speculate that students who, for whatever reason, do not accurately estimate their expected grades may be more likely to suffer disillusionment following the receipt of the actual grade, more likely to misunderstand the content of the course, and less likely to base their evaluations of the instructor on the quality of instruction. Perhaps an increase in both the frequency and the amount of course-related information and feedback given to students at all points throughout the semester may help to decrease the magnitude of students’ rosy grade expectations.

However, from a researcher’s standpoint, such bias in expected grades does not necessarily discount the utility of expected grades as a measure. If research focuses on students’ perceptions, such as satisfaction with grades or satisfaction with the instructors, then expected grades—an inherently subjective index—may prove to be the most valid and useful correlate.

References


Note

Send correspondence to Craig A. Wendorf, Department of Psychology, D241 Science Center, University of Wisconsin, Stevens Point, WI 54481; e-mail: cwendorf@uwsp.edu.

**Students’ Reasons for Writing on Multiple-Choice Examinations**

Frank M. LoSchiavo
Mark A. Schatz
*Ohio University–Zanesville*

Introductory psychology students (N = 142) identified their reasons for writing on a 50-item multiple-choice test and completed a test-taking strategies questionnaire. The majority of students wrote on the exam; we identified 7 distinct reasons. Test performance was significantly related to the total number of writing marks, the total number of reasons identified, and 3 specific writing strategies: highlighting key terms, marking questions that needed further examination, and drawing figures or diagrams. Survey results indicated that the majority of students strongly supported the opportunity to write on tests and that they believed the strategy leads to better performance. We discuss the implications of these results in terms of modern testing formats that preclude the use of writing strategies.

Recognizing the widespread use of objective tests, many universities offer testing skills courses to help students improve their performance on multiple-choice examinations. Through these classes and through various test-taking handbooks (e.g., ACT, 1997; College Entrance Examination Board, 2000), students learn several strategies, some of which require them to write directly on their exams. For example, students learn to highlight important words, eliminate incorrect alternatives, and mark questions that need further review. Given the increasing popularity of computerized testing formats that preclude writing on examinations, it is important to understand why students write on tests to determine if nontraditional evaluation methods are likely to impact performance negatively.

Although researchers have investigated various aspects of test taking, such as answer changing (e.g., Benjamin, Cavell, & Shallenberger, 1984; Shatz & Best, 1987), guessing (Shatz, 1985), and the effectiveness of test-taking skills programs (e.g., Bowering & Wetmore, 1997), only one study has explored the reasons why students write on exams. Kim and Goetz (1993) examined the marks students made on test booklets and determined that students wrote on the test for a variety of reasons. Only one strategy, option elimination, was significantly related to test performance.

The purpose of this study was to further investigate writing strategies on tests. In contrast to Kim and Goetz (1993), students (vs. the researchers) identified their reasons for writing on test booklets. Furthermore, students completed a survey to assess their attitudes toward writing on tests and their general use of the strategy.

**Method**

**Participants**

Participants were 142 students (39 men and 103 women; M age = 21.2 years, SD = 6.09) enrolled in five different introductory psychology classes. All students participated voluntarily for extra credit.

**Procedure**

We collected data from the first of five multiple-choice examinations administered during the term. We administered a similar 50-item multiple-choice test in two sections taught by the first author and in three sections taught by the second author.
Students completed the exam during the first hour of a 145-min class. Written instructions directed students to “use a pencil to record answers on the computerized answer sheet” and to “do as you please with the test booklet.” On completion of the exam, the instructor scored the computerized answer sheets during a 15-min break. The instructor returned the test materials and reviewed the exam on an item-by-item basis.

The instructor then gave students the opportunity to participate in the study (two students declined). The students reviewed their test booklets, and for each instance of writing on the test, they explained their reason on a separate form. Students who did not write on the test explained why they chose not to use the strategy. Afterward, all students completed a brief survey. To assess their attitudes regarding writing on tests, students responded to four items using a 5-point Likert scale. To assess how frequently they wrote on other tests, students responded to one item using the following scale: never, seldom, occasionally, frequently, and always.

Results

Overall, 56% of students wrote on the exam, with women (63%) more likely than men (36%) to do so, $\chi^2(1, N = 142) = 8.49, p < .01$. When listing their reasons for writing, students often used different phrasing to describe the same basic reason. For example, one student may have stated that he or she “crossed out wrong answers,” whereas another student may have stated that he or she “eliminated incorrect options.” To address this issue, we examined a subset of the responses and developed categories for classifying students’ reasons. We identified seven student-generated reasons for writing on the exam. Two coders then worked independently to classify each response. Interrater reliability was high, as coders agreed on 99% of the cases.

On average, writers wrote on 17.95 ($SD = 19.46$) items, made 26.25 ($SD = 33.30$) separate writing marks, and identified 2.73 ($SD = 1.75$) unique reasons. As predicted, the total number of writing marks was significantly related to test performance, $r(140) = .15, p < .05$, one-tailed. Furthermore, performance was significantly related to the number of unique reasons students reported using, $r(140) = .20, p < .05$.

Each reason and its relation with test performance appears in Table 1. Although the effects were relatively small, three reasons were significantly related to performance: highlighting key terms, marking questions that needed further examination, and drawing figures or diagrams. Additional reasons for writing were to indicate correct answers, elaborate on the questions, eliminate incorrect options, and write down additional information (“data dump”).

Of those who chose not to write on the test, most stated that writing was not necessary (76%). Although students could do as they pleased with the test booklet, more than half of the nonwriters stated that they refrained from writing because instructors asked them not to write on exams in the past (52%). Students also stated that they chose not to write because they have become accustomed to it (14%), they never thought of doing it (11%), and they wanted to keep the exam neat (8%).

Survey results indicated that students view writing as a valuable strategy and that they attach great importance to it. The majority of students reported that they write on multiple-choice exams at least occasionally in their other classes (57%), and they either somewhat agreed or strongly agreed that “writing on a test booklet is a valuable test-taking strategy” (72%). Furthermore, they either somewhat agreed or strongly agreed that “students should be allowed to write on a test booklet” (84%), “test performance improves when students are allowed to write on a test booklet” (62%), and “test performance suffers when students are not allowed to write on a test booklet” (47%).

Discussion

Results indicated that writing on test booklets was pervasive, perceived by students as important, and positively associated with test performance. Writing strategies appear to fall in one of two categories. The first category includes strategies that address the mechanics of completing a test, such as indicating correct answers and marking questions that need further examination. The second category includes strategies that help test takers mentally organize complex information, such as eliminating incorrect options, highlighting key terms, drawing figures or diagrams, and writing additional information that is relevant in answering test questions. Correlational analyses indicated that both categories were positively related (albeit weakly) to test performance.

Our findings suggest that the desire to write on a test and the impact writing has on performance may be mediated by test content and test-taker characteristics. For example, 95% of students in an upper level educational psychology class chose to write on their test booklets (Kim & Goetz, 1993), whereas in our sample of introductory psychology students, only 56% chose to write. Although the two studies differed in terms of which specific writing strategies were related to test performance, both studies demonstrated that most students chose to write on test booklets, there was a common set of reasons for writing, and that writing was positively related to test performance. However, given the correlational nature of both studies, we cannot draw firm conclusions about causality. An experimental manipulation would be necessary to determine if writing on a test booklet directly influences test performance.
Because writing on a test booklet appears beneficial and most students attach great importance to it, instructors should carefully consider any testing formats that preclude its use. The most obvious implications involve computerized tests. Although many features make computerized tests attractive (e.g., automatic scoring), test takers may be unable to clerically and cognitively organize the test as they have become accustomed to doing with more traditional testing formats. As Internet-based courses become more popular and computerized testing becomes the norm, experimental research will be necessary to determine if testing formats that preclude writing may negatively influence performance and misrepresent the abilities of test takers who like to write on tests.

References


Notes

1. We reported preliminary findings at the seventh annual American Psychological Society Institute on the Teaching of Psychology in Miami, FL, June 2000.
2. Send correspondence to Frank M. LoSchiavo, Department of Psychology, Ohio University–Zanesville, 1425 Newark Road, Zanesville, OH 43701; e-mail: loschiav@oak.cats.ohiou.edu.

The Teaching of Psychology Course: Prevalence and Content

William Buskist
Rachel S. Tears
Auburn University

Stephen F. Davis
Karen M. Rodrigue
Emporia State University

We present the results of a national survey on the prevalence and content of courses on the teaching of psychology for graduate teaching assistants (GTAs). Ninety-eight (67%) of the psychology departments we surveyed have a formal course on the teaching of psychology. These courses tend to be 1 academic term in length, involve observation of GTAs’ teaching and feedback, and vary moderately in content. Our results prompt several suggestions for designing and implementing new courses on the teaching of psychology or revising extant ones.

Large colleges and universities frequently employ graduate teaching assistants (GTAs) to assist faculty or to serve as teachers of record in introductory level courses. Many of these institutions provide training to their GTAs in an attempt to prepare them for the responsibilities inherent in teaching (e.g., Eckstein, Boice, & Chua-Yap, 1991; Lumsden, Grosslight, Loveland, & Williams, 1988; Mueller, Perlman, McCann, & McFadden, 1997). Such efforts are important because they introduce GTAs to the basic principles of effective teaching, which in turn, may enhance the learning experiences of the undergraduates they teach.

In an extensive study of doctoral-granting schools that include GTA training programs, Meyers and Prieto (2000a) found that most GTAs at these schools received teacher training through the department, another unit of the university, or a combination of these. In some cases, though, GTA participation in the training program is voluntary, and in others GTAs may not participate in all training activities. They also found that the extent of training varied considerably across schools, ranging from none to extensive training.

Some GTA training programs provide formal instruction on teaching (Benassi & Fernald, 1993; Grasha, 1978) and combine GTAs enrolling in a course or seminar on teaching with actual classroom teaching (Rickard, Prentice-Dunn, Rogers, Scogin, & Lyman, 1991). Unfortunately, little is known about the extent and nature of such courses. We conducted a national survey to address this important issue. Our study differed from Meyers and Prieto’s (2000a) research in that we focused exclusively on the teaching of psychology course.

Method

Participants and Procedure

We mailed a cover letter, informed consent information, and questionnaire to chairs of psychology departments at 365 U.S. colleges and universities. We selected these departments based on the availability of teaching assistantships for graduate students as described in the American Psychological Association’s (APA; 1998) Graduate Study in Psychology. We asked department chairs to have the faculty member responsible for supervising or training GTAs complete and return the survey. We did not send a follow-up request to those departments who failed to respond.

A total of 236 (65%) departments responded to the survey; 146 (62%) of these departments offered teacher training for their GTAs. Ninety-eight (67%) of these 146 departments reported offering a course on the teaching of psychology. These 98 departments represent 42% of the 236 departments that responded to the survey. This percentage is almost identical to the 43% reported by Meyers and Prieto (2000a). Twenty-nine respondents indicated that they did not employ GTAs.
15.86, \( p < .0001 \), indicating that quizzes and the cumulative exam were more influential on class attendance (\( M = 2.60, \ SD = .82 \)) than on study behavior (\( M = 2.34, \ SD = .73 \)). There were significant correlations between attendance and course grade, \( r(66) = .40, \ p = .001 \); attendance and quiz performance, \( r(66) = .81, \ p < .0001 \); and quiz performance and course grade, \( r(66) = .63, \ p < .0001 \).

Discussion

The results suggest infrequent and minimal-weight quizzes had a stronger influence on self-reported motivation for attendance and study behavior than a cumulative exam. Consistent with prior research, this study supports the use of unannounced quizzes to improve attendance and study behavior and offers one important distinction. Wilder et al. (2001) used weekly extra-credit quizzes worth up to 6% of course points, and Ruscio (2001) administered more frequent quizzes accounting for 15% of course points. However, this study used only six quizzes with a combined worth of only 3% of course points. Because there is a relation between attendance and course grade, and infrequent and minimal-weight unannounced quizzes influence students to attend class more often, the use of these quizzes can be beneficial to students. Future research can contribute to understanding the relation between unannounced quizzes and preclass preparation by using objective measures of study behavior (e.g., having students record their class-related study time).

Students rated the use of infrequent and minimal-weight quizzes favorably. Students \( (n = 32) \) evaluated a list of course activities and assignments on a scale from 1 \( (\text{among the worst}) \) to 7 \( (\text{among the best}) \). One item asked students to rate the use of unannounced quizzes; the average rating was 5.75. Moreover, 84% of students gave a rating of 5 or higher, indicating that a majority of students rated these quizzes at least above average when compared to other course activities and assignments.

The use of unannounced quizzes has several benefits in addition to improving attendance and study behavior. These quizzes expose students to exam-type questions, help minimize test-taking anxiety, provide immediate feedback on students’ learning, identify areas of content difficulty to the instructor, and potentially improve long-term learning. An added benefit is that with infrequent administration, unannounced quizzes are simple to implement and require minimal time both in and out of the classroom.

References


Notes

1. Special thanks to David Hansen and Thea Rothmann for their outstanding editorial assistance.

2. Send correspondence to Haig Kouyoumdjian, Department of Psychology, University of Nebraska–Lincoln, Lincoln, NE 68588–0308; e-mail: hkouyoum@unlserve.unl.edu.

Differential Test Performance From Differently Colored Paper: White Paper Works Best

Nicholas F. Skinner
King’s College, The University of Western Ontario

Scores on multiple-choice tests printed on colored paper were uniformly lower than scores for tests printed on white paper. These results call into question the efficacy and fairness of the practice of attempting to reduce cheating by printing alternate forms of tests on paper of different colors.

One strategy for controlling cheating is printing tests on differently colored paper (e.g., Sinclair, Soldat, & Mark, 1998). However, studies of the effects of color on test performance have yielded contradictory results. Jacobs and Blandino’s (1992) findings using the Profile of Mood States (McNair, Lorr, & Droppleman, 1971) printed on white, blue, green, red, and yellow paper suggested the possibility of superior performance on tests printed on red and yellow paper because these colors decrease levels of test anxiety. On the other hand, Sinclair et al. (1998) reported better performance on blue paper than red paper in an actual examination setting. They explained this finding in terms of their cognitive tuning.
extension of the affect-as-information hypothesis. Specifically, blue paper may enhance performance by signaling “that the situation is relatively more serious and that systematic processing is more necessary” (p. 131), whereas red paper might impair performance by signaling a more neutral situation less likely to require systematic processing. However, Sinclair et al. did not employ a white paper control condition. In contrast, Soldat, Sinclair, and Mark (1997) included a white paper control group, but measured only the affect conveyed by the paper and not performance. In their second experiment they used red and blue paper but not white paper and measured performance on a simulated rather than actual test.

Previous research thus supports diametrically opposed hypotheses about the relation between color of paper and test performance. Such disparate outcomes call into question the hypotheses about the relation between color of paper and test performance. In their second experiment they used red and blue paper but not white paper and measured performance on a simulated rather than actual test. Previous research thus supports diametrically opposed hypotheses about the relation between color of paper and test performance. Such disparate outcomes call into question the hypotheses about the relation between color of paper and test performance. In their second experiment they used red and blue paper but not white paper and measured performance on a simulated rather than actual test.

Method

After choosing their seats in a tiered classroom, 265 undergraduate students (163 women, 102 men; M age = 19.6 years, range = 17 to 38 years) took a 60-item introductory psychology multiple-choice midterm test. Each test was printed in black ink on blue, green, red, yellow (i.e., the four psychological primary colors), or standard white paper. The questions were identical on all forms of the test and occurred in the same order for all five paper colors. Tests were distributed in the order white–blue–red–green–yellow, and no adjacent individuals received the same color. Students indicated their answers directly on the test papers, which were graded by two independent scorers (interscorer agreement = 100%).

Results and Discussion

A 2 (gender) × 5 (color) ANOVA revealed significance only for the color main effect, F(4, 258) = 7.14, p < .05. According to Tukey post hoc analyses of all possible pairwise comparisons (using a Bonferroni adjustment; p < .004), scores for tests written on green (M = 39.6, SD = 12.0) and blue (M = 40.6, SD = 11.5) paper were significantly lower than scores on yellow (M = 41.5, SD = 11.5) and red (M = 41.9, SD = 10.2) paper, which in turn were significantly lower than scores on white paper (M = 43.9, SD = 9.8).

The absence of a gender effect replicated the findings of Sinclair et al. (1998), as did the demonstration that some colors produced poorer performance than others.1 The most important discovery of this research, which could not be addressed by Sinclair et al. because they did not include a white paper control group, is that the best performance was elicited by white paper. The recommendation by Sinclair et al. urging instructors not to use color-coded forms is now empirically validated. Indeed, these results suggest that the use of colored test paper is counterproductive, raising the question of how colored paper detracts from performance. For example, are these findings attributable to contrast or novelty effects, the influence of affective cues on cognitive processing, or emotional responses triggered by the colored paper (e.g., heightened anxiety according to the Yerkes–Dodson Law)? In addition, the results show clearly that effectively deterring cheating requires a strategy other than using differently colored paper, such as using alternate forms of a test (McBurney, 1999; Whitley & Keith-Spiegel, 2002) printed on white paper only, perhaps combined with assigned seating.

References


Notes

1. I thank Randolph Smith and three anonymous reviewers for their valuable comments on earlier drafts, particularly Reviewer 3 for a number of cogent and collegial suggestions about restructuring portions of the article.

Students who took the test on colored paper consequently received an appropriate upward adjustment of their scores.
The Most Frequently Cited Books in Introductory Texts

Richard A. Griggs
University of Florida

Derrick L. Proctor and Sheila M. Cook
Andrews University

Gorenflo and McConnell (1991) found little commonality among introductory textbooks with respect to journal article citation. Using a stratified random sample of 15 current texts, we extended this research by checking for commonality in book citations. We found little commonality. Of 3,608 unique book citations, 70% were in only 1 text, 91% in 3 or fewer texts, and only 17 met the commonality criterion of citation in at least 80% of the texts. No book was cited in all of the texts, and there was a strong currency citation bias with 70% published since 1980. We discuss these findings in terms of the nonhomogeneity of introductory texts and the larger fractionation problem in contemporary psychology.

Griggs and Marek (2001) systematically reviewed extant research on introductory textbooks and found that the perceived similarity of these texts was a case of stereotyping based on two global text variables: chapter topics and organization. Beyond these dimensions, textbook homogeneity disappeared. For example, analyzing a sample of 24 introductory textbooks published between 1985 and 1989, Gorenflo and McConnell (1991) found 37,590 individual bibliographic entries, but not one journal article was cited in all texts. Only 3 articles were cited in at least 90% of the texts and 14 in at least 75%. These findings indicate that introductory textbook authors use idiosyncratic criteria for deciding which articles to cite.

Surprisingly, Gorenflo and McConnell (1991) also found that the names of many historically prominent psychologists like Freud, Skinner, and Piaget were missing from their list of most frequently cited journal articles. To explain this observation, Gorenflo and McConnell reported that a qualitative analysis of the master bibliographic list suggested that references to these psychologists were primarily to their books and not their research articles. However, Gorenflo and McConnell did not provide these data or any quantitative analysis of these data to support their claims about book citations. Thus, whether eminent psychologists are indeed represented via book citations remains an open question, as does the more important question of whether there is any more commonality among textbooks with respect to references to books versus journal articles. We examined both of these questions in this study. Thus, this study extends the work of both Gorenflo and McConnell and Griggs and Marek (2001).

Given that the books of influential psychologists have more impact than their journal articles (see Heyduk & Fenigstein, 1984; Simonton, 2002), we assumed that their books would be commonly referenced. However, because a few eminent psychologists such as Freud and Piaget have several books that could be cited, different text authors may reference different books, lessening the commonality of citation for these psychologists. In addition, the emphasis on currency of references in introductory texts (Blumenthal, 1990; Weiten & Wight, 1992) may also reduce such commonality as well as book citation commonality in general. Griggs and Mitchell’s (2002) finding that there were few classic core concepts (i.e., core terms appearing in both introductory texts from the 1950s and now) renders this possibility more plausible. If the core language in the texts has changed, then it is certainly reasonable to expect that the book citations have also changed to be more current. In summary, it is likely that the books of historically prominent psychologists are not commonly cited in introductory texts, and, as found for journal article citations, there is little commonality in book citation in general.

Method

In their compendium of introductory psychology textbooks, Koenig and Griggs (2002) listed 38 current (latest copyright from 1999 to 2002) introductory texts, grouped according to 5 levels of difficulty (from low to high). This set of texts included all those currently available for adoption except for the briefer versions of some textbooks. To obtain a stratified random sample of 15 textbooks, we randomly sampled 3 texts from each of the 5 levels. Reference information for the 15 textbooks appears in the Appendix. We examined the reference list in each text for references to books and entered each one into a computer file. We included both authored and edited books. With one exception, we left different volume numbers in a series of books as distinct entries. The exception was the two volumes of James’s (1890) Principles of Psychology, which we counted as one book. We computed the total number of book citations for each text and some overall descriptive statistics: median number of citations for a text and the total numbers of citations and different citations across texts. We then compiled a list of the books cited in all 15 textbooks, those cited in 14 of the 15 texts, and so on. Finally, we computed the number of cited books published before 1900, for each decade of the 20th century, and in 2000 or later. These data allowed us to check for any citation publication-date trends.

Results and Discussion

There were 6,212 total book citations in the reference sections of the 15 textbooks, with individual totals ranging from 187 to 644. The median number of book references per textbook was 424. There were 3,608 different books cited. Table 1 gives the cumulative breakdown for these 3,608 books. The results were much like those of Gorenflo and McConnell (1991) for journal articles in that there was not much commonality. Of the 3,608 books, 70% (2,542) were cited in only 1 text, and 91% (3,272) were cited in 3 or less of the 15 texts. No book was cited in all 15 texts.

Rather than apply unanimity as our criterion for citation commonality, we employed the more liberal 80% criterion
the time they construct the card and thus do not enhance student exam performance. Dickson and Miller (2005) found that most students reported that the amount of time they spent studying did not increase when they constructed a crib card compared to exam preparation without a crib card. Thus, the time spent constructing a crib card may have diverted students from more productive study activities. Some students reported that they did not learn the course material as well when they had a crib card, which suggests that the process of constructing a crib card interfered with learning. In contrast, a minority of students thought that the process of constructing a crib card was a useful study strategy. Crib card construction and use are typically confounded because they are not examined as separate processes. Thus, it is unclear whether creating a crib card is beneficial for student learning and exam performance.

The purpose of this study was to separate the effect of crib card construction from the effect of crib card use on multiple-choice exam performance. Students in an upper division psychology course used self-constructed crib cards for two of the four exams. For the other two exams, they unexpectedly used a crib card constructed by another student.

Method

Thirty-two child and adolescent development students (82.4% women; 14.7% sophomores, 35.3% juniors, 50.0% seniors; 73.5% psychology majors) constructed and used crib cards (one side of a 5 × 8 in. [12.5 × 20 cm] index card) for Exams I and III (self-constructed crib card). The instructor directed students to draw on the course learning objectives and the textbook when constructing the cards. For Exams II and IV, students were not aware that they would be able to use a crib card. As the instructor handed out these exams, she also distributed premade crib cards to students for use during the exams. A student research assistant developed crib cards based on the learning objectives and textbook (other-constructed crib card). The research assistant had neither taken the course nor seen the exams prior to developing the crib cards. The first author reviewed the other-constructed crib card for accuracy and found no errors.

Effect of Crib Card Construction and Use on Exam Performance

K. Laurie Dickson and Michelle D. Miller
Northern Arizona University

Both normal classroom use and research typically confound crib card construction with crib card use, making it unclear whether students benefit from the process of creating crib cards. We compared the effect of self-constructed and other-constructed crib cards (written by a student research assistant) on undergraduates’ (N = 32) multiple-choice exam performance. Performance was better with other-constructed cards than with self-constructed cards. Crib card construction did not facilitate student learning, nor did the use of self-constructed crib cards enhance exam performance.
Students took four multiple-choice question exams (50 questions each) derived from the textbook test bank and based on the learning objectives. Seventy-six test items required lower order skills, and 124 test items involved higher order skills. (For definitions of these terms, see Dickson & Miller, 2005.)

Results

To examine the effect of self-constructed versus other-constructed crib cards on exam performance, we computed each student’s average score for self-constructed crib card exams (M = 70.54, SD = 9.03) and other-constructed crib card exams (M = 76.00, SD = 8.49). Students performed better with the other-constructed cards than with the self-constructed cards, t(31) = 4.16, p < .001.

Discussion

Exam performance was significantly better when students unexpectedly used a crib card constructed by someone else. A likely explanation for this finding is that when students do not expect to use a crib card, they make a stronger effort to understand the material prior to the exam. In contrast, when they expect to have information available on a crib card during the exam, they do not study the material as effectively. Thus, the unexpected crib card contributes to their higher scores by serving as a memory aid or as a resource for information that they did not learn. In other words, students simply use the information on their own crib cards to answer the exam questions without learning the material. It is possible that students actively process test-related information during crib card construction, but even if so, this active processing does not compensate for the overall reduction in learning associated with self-constructed crib cards.

Another possible explanation for our results is that the quality of the other-constructed crib cards was significantly better. This possibility seems unlikely given that a student who had not taken the course constructed the crib cards with the textbook and the learning objectives, just as the students in the class did. Even if the other-constructed cards contained more information, the finding that students performed better with the premade crib cards than with their own contradicts the coding hypothesis (Dorsel & Cundiff, 1979; Hindman, 1980; Trigwell, 1987) because it suggests that having information on the crib card helps more than the process of constructing the crib card.

A related issue is that average students may not know how to effectively engage in the coding process while constructing a crib card. In other words, they may not know what information to put on the card or how to study the information on the crib card, deficits that highlight weaknesses in student study habits in general. The fact that our research assistant, an above-average student, had not taken the course or seen the exams, yet produced a crib card that led to higher exam scores, supports this idea. Future research should examine the coding hypothesis with an intervention study by teaching students how to effectively create and use a crib card.

The construction of crib cards did not facilitate student learning, and the use of self-constructed crib cards did not enhance exam performance in this content-driven upper division psychology course. Given the mounting evidence against crib card effectiveness (Dickson & Miller, 2005; Hindman, 1980; Whitley, 1996), we recommend against the use of crib cards during exams, until teaching scholars more fully understand how crib cards affect learning and exam performance after instructors have coached students on the processes of actively selecting, organizing, and representing the information in a condensed form.

References


Notes

1. We thank the anonymous reviewers for their insightful comments and Carl Moncher for making the other-constructed crib cards.
2. Send correspondence to K. Laurie Dickson, Department of Psychology, Northern Arizona University, P.O. Box 15106, Flagstaff, AZ 86011; e-mail: laurie.dickson@nau.edu.

Effect of Study Guide Exercises on Multiple-Choice Exam Performance in Introductory Psychology

K. Laurie Dickson, Michael S. Devoley, and Michelle D. Miller
Northern Arizona University

We experimentally investigated whether requiring completion of only multiple-choice study guide questions differentially affected multiple-choice exam performance compared to requiring a variety of study guide exercises (learning goals, vocabulary, fill-in-the-blank concepts, multiple-choice questions, short essay questions, matching, and language enhancement) for introductory psychology students (N = 244). There was not a significant difference in multiple-choice exam performance. We conclude that requiring a variety of study guide exercises produces no significant benefit for multiple-choice exam performance over and above those produced by requiring only multiple-choice exercises.

Published study guides that accompany psychology textbooks potentially enhance student learning and exam perfor-
Effect of Paper Color and Question Order on Exam Performance

Ilanit R. Tal, Katherine G. Akers, and Gordon K. Hodge
University of New Mexico

To deter cheating, teachers commonly use exams printed on differently colored paper or with varied question orders. Previous studies, however, reported that paper color and question order affect exam performance and suggested that teachers should adjust students’ scores accordingly and discontinue the use of alternate exam forms. We conducted 2 experiments testing the effects of paper color and question order on exam performance. Students performed worse on exams printed on primary blue paper, but pastel colors had no effect. We found no effect of exam question order (sequential vs. random). We conclude that colored paper or differently ordered questions do not affect exam performance.

Teachers often use alternate exam versions to curb cheating. However, differently colored paper (Sinclair, Soldat, & Mark, 1998; Skinner, 2004; Soldat, Sinclair, & Mark, 1997) and differently ordered questions (Balch, 1989) commonly distinguish alternate versions, but appear to affect performance.

Although Michael and Jones (1955) reported no effect of paper color on exam performance, several more recent studies have contradicted these results. Soldat et al. (1997) found students scored higher on difficult GRE-type questions printed on blue compared to red paper. Sinclair et al. (1998) reported students scored higher on a midterm exam printed on blue versus red paper. These researchers supposed that the color red induces positive affect, triggering a processing strategy that conserves cognitive resources. On the other hand, the authors suggested that the color blue induces negative affect, triggering detail-oriented processing that might be more effective when solving difficult problems. Thus, they argued that better performance on blue versus red exams may result from different affective states and cognitive processes induced by the different colors (cf. Elliot, Maier, Moller, Friedman, & Meinhardt, 2007). In contrast to these findings, Skinner (2004) reported that students scored higher on exams printed on red than on blue paper, with the highest scores occurring on exams printed on white paper. Skinner (2004) and Sinclair and colleagues (1998) urged teachers to discontinue use of differently colored exam versions or at least to adjust students’ exam scores to account for any disadvantages resulting from using different colors of paper.

Concerning question order, students performed better on exams containing questions paralleling the order of presentation in lectures and the textbook (sequential order) than on exams containing randomly ordered questions (Balch, 1989). To explain this effect, Balch suggested that items are best retrieved from memory in the same context in which they were encoded. Exams with sequentially ordered questions present the items in the same order in which they were presumably encoded, and therefore students would perform better under this condition. Other studies, however, have not replicated Balch’s effect (Goss Lucas & Bernstein, 2005; Neely, Springston, & McCann, 1994; Perlini, Lind, & Zumbo, 1998).

In our classes, we use pastel colors to distinguish different exam versions rather than the vivid primary colors used in previous research (N. F. Skinner, personal communication, November 14, 2005; Sinclair et al., 1998; Soldat et al., 1997), and we present multiple-choice questions and alternative answers in random rather than sequential order. Our concern was whether the effects found with primary colors also occur with pastel colors (i.e., red vs. pink). In this study, we attempted to replicate previously reported effects of primary-colored paper and also tested whether pastel colors of paper affect exam performance. To address whether question order (sequential vs. random order)
affects exam performance, we included question order in our study design. In addition, we included gender in our analyses because previous researchers (Elliot et al., 2007; Michael & Jones, 1955; Sinclair et al., 1998) considered student gender as a factor that might interact with paper color to affect exam performance.

Method and Results

Experiment 1: Exams on Primary-Colored Paper

Undergraduate introductory psychology students (364 men, 230 women) took a 40-question multiple-choice exam based on textbook (Davis & Palladino, 2004) and lecture content. Exams were printed in black ink on either white paper (Xerox Business Multipurpose 4200 Bright 92) or four primary colors of paper: RIV 02054 rojo red, RIV 02055 lemon yellow, RIV 02057 emerald green, and RIV 02059 marine blue (Riverside Paper Kaleidoscope Multipurpose). Within each color, questions appeared in either sequential or random order, resulting in a total of 10 exam versions. Students seated adjacently could not have the same color of exam; seating was unassigned.

We performed an ANOVA (SPSS Version 12.0) with number of questions correct as the dependent variable and paper color, question order, and gender as fixed factors. We used Type II sums of squares to correct for unequal cell sizes and tested a priori contrasts based on previous research. We found only a significant main effect of color, $F(4, 594) = 5.53$, $p < .001$, $\eta^2 = .04$, but this model explained only 3.3% of the variance in exam performance (adjusted $R^2$). Contrasts revealed that students performed better on white exams than all other colored exams ($M$ difference $= 2.15$), $t(183.26) = 2.85$, $p < .01$, and worse on blue than all other colored exams ($M$ difference $= -3.27$), $t(197.6) = -4.34$, $p < .001$. We corrected degrees of freedom for unequal variances. We found no effects for question order or gender.

Experiment 2: Exams on Pastel-Colored Paper

A different group of undergraduate introductory psychology students (434 men, 247 women) took a 40-question multiple-choice exam under similar conditions as those described in Experiment 1, except that exams were printed on either white paper (Xerox Business Multipurpose 4200 Bright 92) or four pastel colors of paper: 32571 pink, 32541 [light] yellow, 32561 [light] green, 32521 [light] blue (Wausau Paper Exact Multipurpose).

We performed an ANOVA with number of questions correct as the dependent variable and paper color, question order, and gender as fixed factors. We tested a priori contrasts as in Experiment 1. We used Type II sums of squares to correct for unequal sample sizes and found no effects of color, question order, or gender.

Discussion

We found that printing exams on primary but not pastel colors of paper affected exam performance. Consistent with Skinner (2004), students performed better on exams printed on white paper than on vividly colored paper. Also consistent with Skinner, students performed worse on blue exams compared to other colored exams; this finding, however, contradicted the findings of Sinclair et al. (1998) and Soldat et al. (1997), who found worse performance on red exams. Although the conflicting evidence regarding the directionality of the red and blue paper effects on exam performance may be a result of differences in the hues of paper colors used among studies, the size of our primary color effect was relatively small and toward the lower end of the normal range of .01 to .14 (Cohen, 1988). We could retroactively calculate effect size (Haase, 1983) only for the Sinclair et al. (1998) study, and it was small ($\eta^2 = .02$) as well. Given the small effect sizes and contradictory findings across studies with respect to which colors improve or decrease performance, we conclude there are no meaningful or consistent effects of primary color on exam performance.

We repeated the experiment using pastel colors of paper, which is how we distinguish alternate exam versions in our classes, and found no effect of paper color on exam performance. This lack of effect may occur because lighter colors do not induce the same affective and cognitive states as more vivid colors. Regardless, concerns regarding unfair advantages resulting from different colors of exam paper are alleviated when teachers use pastel rather than primary colors of paper.

In addition to paper color, we also tested whether question order affected exam performance. Although Balch (1989) found that students performed better on exams containing questions in sequential compared to random order, the results of our study as well as those by Goss Lucas and Bernstein (2005), Neely et al. (1994), and Perlini et al. (1998), did not support a question
order effect. Also, consistent with previous studies (Elliot et al., 2007; Michael & Jones, 1955; Sinclair et al., 1998), we found no effects for gender and no interaction with gender.

In summary, we found that paper color and question order do not meaningfully predict differences in exam performance. Faculty can use differently colored paper or differently ordered questions to deter cheating and not worry about affecting grades.

References


Notes

1. Our thanks to Kathryn Wiggins, Harold Delaney, Randolph Smith, and three anonymous reviewers.

2. Send correspondence to Gordon K. Hodge, Department of Psychology, Logan Hall, MSC 03 2220, University of New Mexico, Albuquerque, NM 87131–0001; e-mail: ghodge@unm.edu.
The Exam-A-Day Procedure Improves Performance in Psychology Classes

Frank C. Leeming
University of Memphis

Students in 4 classes had a short exam at the start of every class. Grades were significantly better than in previous classes where the same material was taught but with only 4 exams, and there were fewer withdrawals from the exam-a-day classes. Students taught with the exam-a-day procedure also performed better on a retention test than did students from classes taught using less frequent exams. Responses to anonymous questionnaires administered in all 4 classes indicated that most students believed having an exam every day led to more studying and better learning than in their other classes and that they liked the procedure.

My teaching style for many years included four scheduled exams per semester with an optional final exam for students who wished to drop one grade. Exams consisted primarily of short-essay questions taken from about 75 study questions included in the course syllabus. Each exam also included several fill-in-the-blank or definition questions intended to ensure that students read the textbook.

Although many students in my classes have generally made high grades, I have been unhappy with the number of Ds and Fs. My impression has been that poor performance has most often been the result of simply not enough studying rather than students' lack of ability. Regarding study time, Michael (1991) discussed what he called the "procrastination scallop." He argued that the college environment provides many activities that compete with studying and lead to procrastination until an exam is imminent. Of course, the resulting last minute cramming is often insufficient. Michael proposed more frequent exams as a remedy for the procrastination scallop.

Bangert-Drowns, Kulik, and Kulik (1991) conducted a meta-analysis of 35 studies of the effects of frequency of classroom testing. All studies that measured student attitudes found more frequent tests related to more positive attitudes about the class, and 29 showed a positive effect on final-exam performance. The average effect size (ES) of the 35 studies was .23. Twelve of the studies used psychology classes and 8 of these found a positive effect, although the average ES was only .10. However, no one gave exams more often than once per week, and the average was an exam only every 10.01 days. Furthermore, performance on a scheduled final exam was the only outcome measure analyzed. In these circumstances, any effect of frequent exams might be masked by the tendency of most students, regardless of number of previous exams, to study hard for a final exam.

Michael's (1991) rationale and the data reviewed by Bangert-Drowns et al. (1991) convinced me to use more frequent testing. I chose to give an exam every day, although I had never heard of anyone using this strategy. Results from my first class were such that I have continued to use the technique in subsequent undergraduate classes. The following sections provide a brief description of my class format with an exam-a-day and a description of the outcome in four classes, including both attitudes of the students and data comparing performance of these students to those in my previous classes.

Method

Participants

I originally used the exam-a-day procedure in a Learning and Memory course taught in a 5-week summer term (n = 21). I subsequently used the technique in two sections of Introductory Psychology (n = 143) and again in Learning and Memory the following summer (n = 28). In all apparent respects, students in exam-a-day classes were indistinguishable...
from my students in comparison classes from previous years (i.e., gender, major, year in school).

Class Format

Learning and Memory met 5 days per week with each class lasting 100 min. Introductory Psychology met 2 days per week with classes lasting 75 min. Teaching style and class format were the same as in previous classes in almost every respect. I used the same study questions (with minor alterations each term) and syllabi for each course, the same grading procedure and grade scale (90% = A, 80% = B, etc.), and the lecture mode of teaching. The main difference was that I gave an exam during the first 10 to 15 min of each class, beginning with the second class meeting, rather than only four times during the course.

Most exams had two short-essay questions, taken from the pool of study questions provided in the syllabus, and about five short-answer questions based on material from the text or lecture. After the exam, I spent 2 to 3 min discussing the correct answers and then devoted the remaining time to normal teaching activities.

The syllabus stated that students could drop the three lowest exam scores with no penalty. I adopted this procedure to allow for any accident, illness, or situation preventing attendance. Thus, in a typical class with 22 to 24 exams, the semester grade for students choosing not to take the optional final exam was based on their best 19 to 21 exam scores. Students could choose to take a comprehensive final exam to drop an additional three exam grades, but few students took this option.

Retention Test

Volunteer students from three sections of Introductory Psychology, taught during the same semester by three different instructors, completed a 2 hr retention test near the end of the semester. Participation satisfied the departmental requirement for research participation. Forty-eight participants were from the exam-a-day section, and the other 30 were from sections that had only three exams during the semester. All sections used the same textbook and covered the same chapters in the same order. The retention test contained questions covering the first four chapters of the textbook and associated lecture material. This material had constituted a single test block in two sections, whereas students in the exam-a-day section had taken 13 short tests over the material. At least 6 weeks had passed since any of the three instructors had covered this material and, to minimize rehearsal, participants were not informed of the nature of the experiment. Thus, the test clearly measured long-term retention.

Considerable care was taken to construct a fair test (details of the procedure are available on request). The test contained randomly selected short-essay, multiple-choice, and fill-in-the-blank questions from the class exams previously administered by each of the three instructors. There were six separate test sessions and each included students from all of the classes. Instructions on the exam booklet encouraged students to do their best but also stated that scores would be confidential and would have no influence on course grades. Individuals unaware of student identification scored all tests.

Results

Student Reactions

Students in each class completed an anonymous questionnaire about the exam-a-day procedure near the end of the term. Students indicated whether they agreed, disagreed, or were unsure about statements concerning the procedure. Table 1 summarizes responses with data presented separately for the different courses.

Responses consistently showed a positive evaluation of the exam-a-day procedure, and there were no systematic differences between courses. The vast majority of students reported that the procedure had led to more studying, they kept up with the material better, they learned more than in most of their other classes, and they liked the procedure.

Attendance and punctuality improved dramatically over previous classes. This result, of course, was not surprising given that every class began with a timed exam.

Semester Grades

I compared the final semester grades of students in the exam-a-day classes with those of students in the most recent

Table 1. Responses to Survey by Students in Two Courses

<table>
<thead>
<tr>
<th>Statement</th>
<th>Learning and Memory</th>
<th>Introductory Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I was skeptical when the procedure was first described.</td>
<td>48</td>
<td>36</td>
</tr>
<tr>
<td>2. I studied more for this class than for most.</td>
<td>70</td>
<td>9</td>
</tr>
<tr>
<td>3. I learned more than I would have with just a few exams.</td>
<td>79</td>
<td>9</td>
</tr>
<tr>
<td>4. I kept up better than in most classes.</td>
<td>91</td>
<td>3</td>
</tr>
<tr>
<td>5. Given a choice, I would choose this procedure over just a few exams.</td>
<td>67</td>
<td>12</td>
</tr>
<tr>
<td>6. Having this course will help me manage study time in the future.</td>
<td>74</td>
<td>6</td>
</tr>
<tr>
<td>7. I like the test-a-day procedure better than just a few tests.</td>
<td>77</td>
<td>6</td>
</tr>
<tr>
<td>8. Test-a-day has been an awful experience.</td>
<td>6</td>
<td>90</td>
</tr>
<tr>
<td>9. I recommend using the test-a-day procedure next semester.</td>
<td>81</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. The values represent mean percentages of responses over two classes within each course. Percentages do not sum to 100 because responses to the option "unsure" have been omitted.

\(^a_{n = 33}^b_{n = 99}\)
sections of the same classes that I had previously taught with only four exams. In Learning and Memory, the respective means were 89% (SD = 8.13) versus 81% (SD = 17.61). This difference, corrected for unequal variances, was significant, \( t(81) = 2.88, p = .005 \). The ES, calculated using the method of Shadish, Robinson, and Lu (1997), was .64. In Introductory Psychology, the respective means were 80% (SD = 22.92) versus 74% (SD = 22.11). This difference was also significant, \( t(232) = 2.03, p = .04, ES = .27 \).

In Learning and Memory, the percentage of Ds and Fs dropped from 21% in two previous classes to 2% in the two classes with an exam every day, whereas the percentage of As and Bs increased from 67% to 81%. In Introductory Psychology, the percentage of Fs dropped from 27% in two previous classes to 12% in the two classes with an exam every day, whereas the percentage of As increased from 22% to 43%. Percentage of withdrawals remained at 11% in Introductory Psychology but dropped from 21% to 13% in Learning and Memory.

**Retention Test**

Class exam scores were higher for students in the exam-a-day class than for students in the two traditional classes, but analyses found no evidence that the retention test involved a biased sample of students. On the retention test, students from the exam-a-day class scored significantly higher (\( M = 54\% \), \( SD = 13.26 \)) than did those from the traditional classes (\( M = 48\% \), \( SD = 13.34 \)), \( t(77) = 2.05, p = .04, ES = .47 \). An ANCOVA on the retention scores, correcting for differences on the in-class scores, failed to show a significant difference. Thus, the better retention scores of the exam-a-day students was the result of better initial learning rather than any effect on memory processes.

**Discussion**

Despite the work of authors such as Bangert-Drowns et al. (1991) and Michael (1991), I am not aware of anyone in psychology testing students very frequently. However, after repeatedly using the exam-a-day procedure in two different courses, I am convinced that this is a better teaching technique than giving infrequent exams over large amounts of material. Average semester percentages in two courses were more than half a letter grade higher than when I taught the same courses with only four exams. Furthermore, the number of high grades was substantially increased and the number of low grades was substantially decreased in all classes with an exam-a-day. The ESs of all comparisons were well above the average of the 12 studies with psychology classes with an average of an exam every 10 days reviewed by Bangert-Drowns et al. Finally, most students reported that the exam-a-day technique increased studying, increased learning, forced them to keep up with the material, and that they liked the procedure.

I have encountered several objections to the exam-a-day procedure. Most serious is the possibility that the procedure results in superficial learning and little retention. I believe that answering the essay and short-answer questions that constitute my exams requires understanding of concepts and principles at an appropriate level. For example, a question from Introductory Psychology asked “Explain how a nerve impulse (action potential) occurs and how it is an all-or-nothing event.” A question from Learning and Memory asked “Describe Rescorla’s objections to commonly used controls in classical conditioning, and the control that he advocates. Thoroughly explain the rationale for his position.” Students given daily exams have performed better than those tested infrequently on such questions, convincing me that level of conceptual understanding of the material is, if anything, improved by more frequent testing. Regarding retention, students from the exam-a-day class had higher absolute scores than students from the traditional classes on the retention test. Although this difference is due to better original learning, the fact remains that these students knew more of the material than students from traditional classes well after class coverage of that material.

Another objection to the exam-a-day format is that it leaves less time for teaching. In fact, it was necessary for me to eliminate some material from my usual lectures the first time the procedure was used in both courses, but the major effect was to reduce redundancy rather than to omit basic material. I believe that, if anything, my lectures are now better and more focused than they were when exams were widely spaced. In fact, I suspect that many instructors show the same procrastination scallop for covering large amounts of material as students do for studying. The bottom line, however, is amount learned, and all of the evidence that I see indicates better learning by more students with the exam-a-day procedure.

A final objection that deserves mention is work load for the instructor. My two Learning and Memory classes met every weekday for 5 weeks and had an average enrollment of 24 students. I spent about an hour grading papers every day and did not find the time to be onerous, especially given the unusually high quality of the student work. However, if grading should become too time consuming, I would probably try more objective and fewer essay questions. I would be extremely reluctant to reduce exam frequency.

**References**


**Notes**

1. I thank Beverly Schaeffer for her help with the conduct of the retention test.
2. Send correspondence to Frank C. Leeming, Department of Psychology, University of Memphis, Memphis, TN 38152; e-mail: f.leeming@mail.psyyc.memphis.edu.
Elaborations of Introductory Psychology Terms: Effects on Test Performance and Subjective Ratings

William R. Balch
The Pennsylvania State University, Altoona

Undergraduate students participated in an experiment designed to evaluate different types of elaborations on definitions of 16 psychology terms. First, participants received booklets presenting the definition of each term, followed by 1 of several elaborations: an example, a mnemonic, a paraphrase, or a repeated definition (the nonelaborating control condition). Then students received a multiple-choice test consisting of questions both on the definitions and on novel examples of the terms. Compared to repeated definitions, examples and mnemonics—but not paraphrases—improved scores on each type of question. However, students' subjective ratings did not always reflect the effectiveness of the elaborations in improving test performance.

Students generally rate a number of textbook study aids (e.g., boldfaced terms, chapter glossaries, chapter summaries, self-tests) as educationally valuable (Gurung, 2003; Marek, Griggs, & Christopher, 1999; Weiten, Guadagno, & Beck, 1996). Yet there may be little or no relation between reported use of such aids and performance in the course. Correlations between students' grades and their ratings of how much they used various textbook aids are either nonsignificant (Gurung, 2003; Weiten et al., 1996), or significant but small, rs ≤ .17 (Weiten, Deguara, Rehmke, & Sewell, 1999). Balch (2001) obtained similar results with study tips that the instructor provided to introductory psychology students (e.g., self-help quizzes, lecture notes, course performance charts). For five out of six tips, reported use and course grades were not significantly correlated.

Compared to these correlations, experimental manipulation of structured study may produce results that more clearly show an effect of study aids on performance. In one experiment, students read textbook passages that either included or omitted marginal, one-sentence inserts of key concepts (Nevid & Lampmann, 2003). For passages that included the inserts, students' performance on a subsequent content quiz was higher than for passages without the inserts. In another experiment, students who took a practice final exam in introductory psychology did better on the final exam itself than students who reviewed the questions on the practice final without actually taking it (Balch, 1998).

I evaluated several aids for studying psychology terms and based these aids on the principle of elaboration (e.g., Motes & Wiegmann, 1999, pp. 63–64). According to this principle, people's retrieval of information from memory depends on how they elaborated on the information during encoding. In the aforementioned research on study tips, the only tip that produced a significant (albeit small) correlation with course performance was a tip on elaboration (Balch, 2001). However, experiments in which elaboration facilitated learning
under controlled conditions would provide stronger evidence for its effectiveness.

One type of elaboration that appears to facilitate the learning of psychology terms is a mnemonic technique, in which a concrete word that sounds like all or part of a given term is associated with the definition of that term in a mental image. For example, students who learned physiological psychology terms with this technique performed better when tested on both definitions and applications of those terms, compared to students who learned by rote repetition (Carney & Levin, 1998).

In addition to evaluating this mnemonic technique in this experiment, I evaluated two other types of elaboration: the example and the paraphrase. The results of several experiments suggest that these methods facilitate the learning of various materials other than psychology terms. For instance, examples or instantiations of statements improved recall of target words (Stein, Morris, & Bransford, 1978) as well problem solving (Chen & Daehler, 2000). Moreover, paraphrasing of material such as a textbook passage (Friedman & Rickards, 1981), written directions (Glover, Harvey, & Corkill, 1988), and proverbs (Honeck, 1973) facilitated memory of that material.

To date, no researchers have assessed paraphrases or examples in terms of their effect on learning psychology terms. However, students' subjective reactions to paraphrase exercises are positive. For instance, Merwin (2003) taught psychology terms using a paraphrase technique that she adapted from the board game Taboo. In their ratings, 71% of her psychology students agreed or strongly agreed with the statement that this technique was helpful for learning course concepts.

Each of the aforementioned elaborations has the potential to be effective in teaching material relevant to students enrolled in introductory psychology, a course in which students typically must learn and apply numerous terms. Thus, my hypothesis was that each elaboration would improve students' learning of psychology terms. To test this hypothesis, I presented students with definitions of various terms, along with one of the following elaborations on each definition: an example, a mnemonic, a paraphrase, or a repeated definition (the nonelaborating control condition). Then students received a test both on the definitions and on novel examples of the terms. In addition, students rated the study materials on comprehension and the methods of elaboration on helpfulness. The rationale for this approach was to assess how instructor might best elaborate on the psychology terms that they teach (e.g., in lectures or handouts) to improve their students' course performance and subjective perceptions of that material.

Method

Participants

One hundred thirty-six undergraduate introductory psychology students at The Pennsylvania State University, Altoona, participated at the beginning of the Fall 2003 semester. The experiment occurred before the presentation (in class) of any of the psychology terms included in the experiment. The sample consisted of 94 women and 42 men, whose ages ranged from 18 to 55 years (M = 20.15, SD = 5.41). All the students volunteered their participation for a small amount of extra credit.

Materials

Study booklets. In all booklets, the definitions of 16 psychology terms appeared in the same random order. Four of the terms related to each of four different introductory psychology topics: experimental method (independent variable, dependent variable, control condition, confounding variable), areas in the brain (medulla, hypothalamus, thalamus, reticular formation), psychological disorders (dissociative, somatoform, schizophrenic, bipolar disorders), and psychotherapies (client-centered therapy, systematic desensitization, psychoanalysis, rational-emotive therapy). The definitions were similar to those listed in the introductory psychology textbook that I use (Bernstein, Penner, Clarke-Stewart, & Roy, 2003), except that all were about the same length, never referred to other related terms, and were expressed as complete sentences (e.g., “A dissociative disorder is a psychological condition in which a person experiences disruptions in memory, consciousness, or identity”).

In every study booklet, two consecutive pages were devoted to each term. The definition of a given term appeared on the first page, followed by a page on which either an elaboration or a repeated definition appeared. Four different types of booklets constituted the four elaboration conditions, as follows.

Paraphrase booklets presented a different, less formal, and more familiar wording of each original definition (e.g., “A dissociative disorder is a psychological problem in which people may not remember certain events that happen to them or who they are”). Example booklets presented a specific example (e.g., “A person with a dissociative disorder might be found standing by a road somewhere without knowing how he got there, what his name is, or where he lives”). Mnemonic booklets presented a word that sounded like part, or all, of a given term and that was associated in an image with some of the words in the original definition (e.g., “The first two syllables in dissociative sound like disco. Discos are usually dark, and people with a dissociative disorder are in the dark because of disruptions in their memory, consciousness, or identity”). Finally, in repeated definition (control) booklets, there was no elaboration because each definition was followed by an exact repetition of that definition.

On each page of every booklet, students indicated a comprehension rating on a scale from 0 (lowest) to 10 (highest) that appeared below each definition and below each elaboration or repeated definition that followed on the next page. Each of the 32 ratings (2 for each of the 16 terms) reflected perceived understanding of a given term, based on the information provided on that page. These ratings were intended to help students maintain their attention as they studied the booklets and to provide a subjective measure of their comprehension.

Test booklets. All test booklets consisted of the same 32 multiple-choice questions. To simulate a real test, I prepared two forms (A and B) in which the questions appeared in one of two different random orders. Sixteen were definition questions, beginning with “The definition of [a given term] is …”
and followed by four choices. The correct choice was the definition of that term as it had been stated in the study booklets, whereas the incorrect choices were the definitions of the three related terms. On the definition question for dissociative disorder, for instance, the definition of that term—along with the definitions of the three other psychological disorders—were the four choices.

The other 16 questions were example questions, beginning with “An example of [a given term] would be . . . “ All examples in the test questions were different from those in the study booklets that used examples. For the term dissociative disorder, for instance, the correct example was “a person who has frequent blackout periods during which she can’t remember what she did.” The three incorrect choices were new examples of the other three psychological disorders.

On the last page of every booklet, each student completed a single helpfulness rating on a scale from 0 (lowest) to 10 (highest). This rating reflected how much each student believed that his or her particular elaboration (or repeated definition) condition had helped on the multiple-choice test, compared to how he or she might have done if studying only the original definitions.

Procedure

In conducting this experiment with groups of students ranging from 20 to 44 in number, I preassigned each student randomly to one of four conditions: example, mnemonic, paraphrase, or repeated definition. Within each group session, approximately equal numbers of students received each type of study booklet.

I instructed students that for each term in the study booklets, the first page relating to that term would present the definition and the next page would contain additional information about the term: either a repetition of the original definition or else some other type of information about the term. In either case, I explained, the additional information was intended to help them learn the term better. There would be a test on the terms afterward, I told them, but I did not specify the nature of the test.

Students studied each numbered page of their designated booklet until I gave the instruction, “Rate,” at which time they completed the comprehension rating for that page. When I said, “Turn to page [whatever page number was next],” they turned the page, studied the information on that page until I gave the next instruction to rate, and so on. In pacing students through their booklets, I allowed 30 sec for each page: 20 sec for study and 10 sec for completing the comprehension rating on that page.

When students finished the study booklets, I collected these booklets and distributed the test booklets (equally divided between Form A and Form B, the two test-question orders, for each elaboration condition). The interval between students’ finishing their study booklets and their starting the test was about 5 min.

In the test instructions, I explained to students that all test questions were multiple choice. On some of the questions, they were to indicate the definition of a previously studied term; on others, they were to indicate an appropriate example of the term. While on a given page, they were to read the multiple-choice question on that page, answer it without waiting for a separate instruction, and then wait until I instructed them to turn the page. I allowed them 40 sec per page. When they finished all the multiple-choice questions, I instructed them to turn to the last page of the booklet and complete the aforementioned helpfulness rating.

Results

To analyze the main effects of elaboration on the performance measures (definition and example scores) as well as on the subjective ratings (helpfulness and comprehension), I performed separate one-way, between-groups ANOVAs on the elaboration conditions: example, mnemonic, paraphrase, and repeated definition (control). For ANOVAs with significant main effects, I further performed three pair-wise Bonferroni t tests on the condition means (df = 132, with the α level adjusted to α/3 for three comparisons). These planned comparisons—one between each of the elaborations (example, mnemonic, and paraphrase) and the repeated-definition control—tested the aforementioned hypothesis that each elaboration would increase performance measures and subjective ratings, compared to performance and ratings obtained with repeated definitions (see Table 1). I report additional, unplanned comparisons as single (df = 66, unadjusted α level) rather than multiple t tests.

Performance Measures

Table 1 includes the means and standard deviations of both the definition and example scores corresponding to each elaboration condition. Each score was the percentage of correct answers to 16 of the 32 questions on the multiple-choice test.

Definition scores. As the table illustrates, definition scores depended significantly on elaboration condition, F(3, 132) = 4.30, p < .01, η² = .09. The planned comparisons revealed that examples (76.63%) and mnemonics (77.94%) significantly improved definition scores compared to the repeated-definition control (65.25%), ps < .05. However, paraphrases (70.56%) were not significantly more effective than repeated definitions, p > .10. Additional comparisons between the conditions revealed that examples and mnemonics each produced significantly or almost significantly higher scores than paraphrases, p < .05 and p < .10, respectively. However, there was no significant difference in scores between examples and mnemonics, p > .10.

Example scores. Although these scores tended to be lower than definition scores, a similar pattern of results emerged. Again, elaboration condition significantly influenced example scores, F(3, 132) = 5.22, p < .01, η² = .11. Scores were significantly higher for the same two elaborations—examples (73.38%) and mnemonics (68.75%), p < .01 and p < .05, respectively—compared to repeated definitions (57.69%). For paraphrases (62.31%), scores did not significantly exceed those for repeated definitions, p > .10. Additional comparisons showed significantly higher scores for
examples than for paraphrases, $p < .05$, but no further significant differences between conditions, $p > .10$.

**Subjective Ratings**

_Helpfulness ratings._ As with the performance measures, helpfulness ratings depended significantly on elaboration condition, $F(3, 132) = 11.76$, $p < .001$, $\eta^2 = .21$ (see Table 1). Ratings for examples and mnemonics followed a similar pattern to those of the performance measures. On a scale from 0 (lowest) to 10 (highest), students rated both examples (7.96) and mnemonics (7.85) significantly more helpful than repeated definitions (5.38), $p < .01$. Ratings for paraphrases (7.16) were also significantly higher than those for repeated definitions, $p < .01$, although test scores for paraphrases did not significantly differ from those for repeated definitions (see the “Performance Measures” section). Additional comparisons revealed that students rated examples higher than paraphrases, $p < .05$, but no further differences were significant, $p > .10$.

_Comprehension ratings._ For each student, I calculated two indexes of perceived comprehension. The first is the original-definition rating (the average rating on the first study booklet page devoted to each of the 16 terms, on which the definition of that term always appeared). These ratings, based on a scale from 0 (lowest) to 10 (highest) and ranging from 7.33 to 7.84, were about the same for all elaboration conditions. Thus there was no significant effect of elaboration condition on the original-definition ratings, $F(3, 132) = 1.04$, $p > .10$. A significant effect would be unlikely, because participants in all conditions rated exactly the same definitions.

The second index (see Table 1) is the increase in comprehension rating: the average increase in each student’s 0 (lowest) to 10 (highest) rating of each elaboration or repeated-definition rating of a given term relative to his or her original-definition rating of the same term. For this index, elaboration did significantly influence the amount of increase, $F(3, 132) = 5.01$, $p < .01$, $\eta^2 = .10$. Planned comparisons between each elaboration and the repeated-definition control showed that paraphrases (1.12) increased rated comprehension significantly more than repeated definitions did (0.69), $p < .05$. However, the same was not true of examples (0.88) or mnemonics (0.42), $p > .10$. Thus, compared to repeated definitions, paraphrases improved students’ perceived understanding of the terms but not their test scores (see the “Performance Measures” section). Additional comparisons showed that paraphrases and examples increased comprehension ratings significantly more than mnemonics did, $p < .01$ and $p < .05$, respectively. However, paraphrases did not increase ratings significantly more than examples did, $p > .10$.

**Discussion**

Considering the generally low (Weiten et al., 1999) or nonsignificant (Gurung, 2003; Weiten et al., 1996) correlations between students’ grades and their reported use of various study aids, these aids may have limited value when students are studying on their own. Yet students may benefit more when their instructors provide them with appropriate information augmenting the material to be learned. The results of this study suggest some effective ways for teachers to help their students by using methods of elaboration. When I provided students with either examples or mnemonics to supplement the definitions of basic psychology terms, their test scores on both the definitions and on novel examples surpassed those of students receiving only repetitions of the same definitions. Thus, for examples and mnemonics, the hypothesis that elaborations improve the learning of introductory psychology terms received support. Moreover, students rated these elaborations as more helpful to their test performance than repeated definitions.

Compared to research on mnemonics, the use of examples to improve learning has been less investigated. Although there is some evidence that exemplification enhances problem solving (Chen & Daehler, 2000) and memory of target words in sentences (Motes & Wiegmann, 1999, pp. 63–64; Stein et al., 1978), these results show that examples effectively promote the learning of psychology terms. Therefore, this finding extends the previous literature on study aids and should encourage instructors to increase, or at least maintain, their current use of examples.

My finding that the use of mnemonics is an effective elaboration reinforces previous recommendations that teachers in-

---

**Table 1. Effects of Elaboration on Performance Measures and Subjective Ratings**

<table>
<thead>
<tr>
<th>Elaboration Condition</th>
<th>Definition Score</th>
<th>Example Score</th>
<th>Helpfulness</th>
<th>Increase in Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>$M$ 79.63, $SD$ 16.31</td>
<td>$M$ 73.38, $SD$ 19.06</td>
<td>$M$ 7.96, $SD$ 1.74</td>
<td>$M$ 0.88, $SD$ 0.73</td>
</tr>
<tr>
<td>Mnemonic</td>
<td>$M$ 77.94, $SD$ 17.50</td>
<td>$M$ 68.75, $SD$ 17.13</td>
<td>$M$ 7.85, $SD$ 2.06</td>
<td>$M$ 0.42, $SD$ 1.04</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>$M$ 70.56, $SD$ 16.69</td>
<td>$M$ 62.31, $SD$ 17.00</td>
<td>$M$ 7.16, $SD$ 1.42</td>
<td>$M$ 1.12, $SD$ 0.62</td>
</tr>
<tr>
<td>Repeated definition</td>
<td>$M$ 65.25, $SD$ 23.56</td>
<td>$M$ 57.69, $SD$ 17.13</td>
<td>$M$ 5.38, $SD$ 2.66</td>
<td>$M$ 0.69, $SD$ 0.60</td>
</tr>
</tbody>
</table>

*Based on percentage of correct answers to a 16-question subset of the multiple-choice test. **Subjective rating (after taking the test) based on a scale ranging from 0 (lowest) to 10 (highest). $^a$Average increase in each participant's 0 (lowest) to 10 (highest) rating of each elaboration or repeated definition relative to his or her rating of the original definition rating of the same term. $^b$Different participants in each condition. *$p < .05$, **$p < .01$. Based on three Bonferroni comparisons in which each of the elaborations (example, mnemonic, and paraphrase) is compared with the repeated-definition control condition ($\alpha$ level adjusted for three $t$ tests).
struct students in applying mnemonics to daily life (Shimamura, 1984), to remembering word lists (Schoen, 1996), and to learning names (Smith, 1985) or course material (Carney, Levin, & Levin, 1994). In particular, these results support the finding that mnemonics improve test performance on definitions and applications of psychology terms (Carney & Levin, 1998).

Given these findings, instructors might consider fortifying their existing lectures and handouts with both examples and mnemonics. This strategy would allow teachers some flexibility in how they elaborate on their course material. In writing the material for the examples and mnemonics in this experiment, I found that for some terms, examples occurred to me more readily than mnemonics. For other terms, however, mnemonics seemed like obvious and natural learning aids. For instance, the hypothalamus is an area of the brain that controls eating and drinking. Moreover, it also sounds like hippopotamus, an animal that eats and drinks in large quantities. In practice, drawing on both types of elaboration might add interest and variety to lectures.

Paraphrases did not facilitate test performance, although previous research has demonstrated the effectiveness of paraphrase questions (Friedman & Rickards, 1981), paraphrasing by students (Glover et al., 1988), or paraphrases presented by the experimenter (Honeck, 1973), in learning other types of material. Thus there was no support for my hypothesis that paraphrases were one of the elaborations that would increase the learning of psychology terms. Compared to examples and mnemonics, however, paraphrases may be less expansive elaborations that clarify, more than extend, the meaning of a definition. Yet my participants rated paraphrases higher than repeated definitions with respect both to helpfulness and to increases in their comprehension ratings of the terms. This finding is consistent with the high subjective ratings that Merwin's (2003) students gave to her technique of having them paraphrase definitions of psychology terms. In the method used here, however, I presented elaborations in booklets and students did not create their own paraphrases. The inconsistency between my participants' subjective ratings of the paraphrases as compared with their test scores in this condition suggests that students may not always be accurate judges of whether a given method of elaboration is effective. Therefore, in evaluating the quality of teaching techniques, instructors need to consider empirical outcomes rather than relying only on students' reactions.

My reason for testing elaborations that I provided for students was to simulate how teachers might incorporate such information, briefly and with little extra preparation, into their existing lectures and handouts—formats widely used in introductory psychology. I acknowledge that some instructors might regard providing ready-made elaborations to their classes as precluding independent thought on the part of their students. Yet the finding that both examples and mnemonics improved students' performance not only on definitions of terms, but on recognizing new examples as well, might lessen this concern. In addition, instructors might consider developing separate exercises to instigate and monitor their students' efforts to elaborate on psychology terms. Although this approach might not always be practical in larger, more lecture-oriented classes, it would encourage students to take more responsibility for their learning.

Information presented by instructors to their classes is not a substitute for students' studying on their own. As Table 1 shows, multiple-choice performance in the two most effective conditions in this experiment—examples and mnemonics—averaged only about 79% for definitions and 71% for new examples, with just a 5-min interval between the end of the study period and the beginning of the test. Real tests usually require the retention of material over considerably more time. Nevertheless, when students initially unfamiliar with the material devoted a total of only 1 min of study for each term, these elaborations improved test scores relative to the repeated-definition control. Thus, providing examples and mnemonics to students appears to be an efficient strategy for improving their learning of basic terms in introductory psychology. Whether these elaborations remain effective over long-term retention intervals, however, is an appropriate question for further research.

References


Notes

1. A faculty development grant from The Pennsylvania State University, Altoona, supported this research.
2. I thank Sarah Harshman and Christina Lang for their assistance in conducting and scoring the raw data from the experimental sessions in this study.
3. Send correspondence to William R. Balch, Department of Psychology, The Pennsylvania State University, 3000 Ivyside Park, Altoona, PA 16601; e-mail: wrb3@psu.edu.
Influence of Unannounced Quizzes and Cumulative Exam on Attendance and Study Behavior

Haig Kouyoumdjian
University of Nebraska–Lincoln

Instructors often use grade-related contingencies to motivate students to attend and prepare for class. This study compared the influence of infrequent and minimal-weight unannounced quizzes to a cumulative exam on students’ self-reported motivation to attend class and regularly keep up with class material. Results indicated that the unannounced quizzes had more influence on self-reported motivation to attend class and study regularly than a cumulative exam. Students rated the use of these quizzes favorably, and there are several benefits for their use in addition to motivating students to better attend and prepare for class.

Instructors encourage students to attend and prepare for class in many ways, including making lectures engaging, encouraging class discussion, requiring attendance, and administering quizzes. Increasing attendance and preclass preparation is challenging, as other activities compete with students’ time (e.g., employment, family responsibilities, social activities). Nevertheless, increasing students’ course involvement is important, and studies have reported a positive association between attendance and performance in college courses (Devadoss & Foltz, 1996; Rodgers, 2001; Romer, 1993; Van Blerkom, 1992).

Variables that motivate students to devote time to classes include extracurricular activities, liking of the instructor, and interest in the subject matter (Galichon & Friedman, 1985). Although the instructor has little, if any, control over these variables, the instructor has considerable control over the course grade, which can be an effective motivator for students (Michael, 1991). Researchers have reported that classroom procedures having grade-related contingencies are effective in increasing attendance (Beaulieu, 1984; Hovell, Williams, & Semb, 1979; Wilder, Flood, & Stromsnes, 2001) and preclass preparation (Connor-Greene, 2000; Marchant, 2002; Ruscio, 2001; Sappington, Kinsey, & Munsayac, 2002). The use of unannounced quizzes is especially effective because it is an application of intermittent reinforcement.

The main risk associated with unannounced quizzes is that students may negatively evaluate their use (Connor-Greene, 2000). To minimize the punitive aspect of pop quizzes, Thorne (2000) recommended using them as extra-credit exercises. Others prefer to have pop quizzes as part of their overall exam grades (Burchfield & Sappington, 2000; Ruscio, 2001). Ruscio showed that frequently administering random quizzes accounting for a total of 15% of the course grade increased students’ reading compliance. Unfortunately, Ruscio did not report students’ evaluation of this assessment method.

Although unannounced quizzes have been reported to increase attendance and preclass preparation, researchers have not examined the effects of a cumulative exam on attendance and study behavior in college students. Given the influence grades have on students’ attendance and study behavior, course requirements contributing to a larger percentage of students’ grades may be most influential in increasing course involvement. The purpose of this study was to compare the influence of infrequent and minimal-weight unannounced quizzes to a cumulative exam on students’ self-reported motivation to attend class and keep up with assigned class material.

Method

Participants

Participants included 66 students enrolled in two introductory psychology courses at a large public university. I taught both courses, and each met for 28 class periods. There were 29 male (44%) and 37 female (56%) participants, with 54% freshmen, 23% sophomores, 15% juniors, 5% seniors, and 3% other.

Materials and Procedure

On the first day of class, I told students there would be unannounced quizzes (three multiple-choice questions each) during the semester and that makeup opportunities would not be permitted. However, the total possible points were 18 and only 15 (3% of total course points) were required. Therefore, if students missed one quiz, it did not count against them, and those present for all quizzes could earn extra-credit points. I also told students there would be a 100-point cumulative exam (20% of total course points), which was worth more than any other course requirement.

At the end of the semester, students voluntarily rated statements about the influence that quizzes and a cumulative exam had on their attendance and study behavior (i.e., keeping up with assigned class material). I verbally communicated confidentiality of responses to the students as a group. Students rated each of the following statements on a 4-point scale ranging from 4 (describes me extremely well), 3 (describes me well), 2 (somewhat describes me), to 1 (does not describe me at all): (a) I think having a cumulative final influenced me to attend class lectures more often, (b) I believe having quizzes influenced me to attend class lectures more often, (c) I think having a cumulative final influenced me to better keep up with my reading assignments and class materials, and (d) I believe having quizzes influenced me to better keep up with my reading assignments and class materials.

Results

A 2 (exam, quiz) × 2 (attendance, study behavior) repeated measures ANOVA revealed no interaction of quizzes and cumulative exam as they related to attendance and study behavior. There was a main effect for quizzes, \( F(1, 63) = 23.99, p < .0001 \), indicating that students perceived quizzes as more influential on study behavior and attendance (M = 2.71, SD = .83) than a cumulative exam (M = 2.12, SD = .85). There was also a main effect for attendance \( F(1, 63) = \)
15.86, \( p < .0001 \), indicating that quizzes and the cumulative exam were more influential on class attendance (\( M = 2.60, SD = .82 \)) than on study behavior (\( M = 2.34, SD = .73 \)). There were significant correlations between attendance and course grade, \( r(66) = .40, p = .001 \); attendance and quiz performance, \( r(66) = .81, p < .0001 \); and quiz performance and course grade, \( r(66) = .63, p < .0001 \).

Discussion

The results suggest infrequent and minimal-weight quizzes had a stronger influence on self-reported motivation for attendance and study behavior than a cumulative exam. Consistent with prior research, this study supports the use of unannounced quizzes to improve attendance and study behavior and offers one important distinction. Wilder et al. (2001) used weekly extra-credit quizzes worth up to 6% of course points, and Ruscio (2001) administered more frequent quizzes accounting for 15% of course points. However, this study used only six quizzes with a combined worth of only 3% of course points. Because there is a relation between attendance and course grade, and infrequent and minimal-weight unannounced quizzes influence students to attend class more often, the use of these quizzes can be beneficial to students. Future research can contribute to understanding the relation between unannounced quizzes and preclass preparation by using objective measures of study behavior (e.g., having students record their class-related study time).

Students rated the use of infrequent and minimal-weight quizzes favorably. Students (\( n = 32 \)) evaluated a list of course activities and assignments on a scale from 1 (among the worst) to 7 (among the best). One item asked students to rate the use of unannounced quizzes; the average rating was 5.75. Moreover, 84% of students gave a rating of 5 or higher, indicating that a majority of students rated these quizzes at least above average when compared to other course activities and assignments.

The use of unannounced quizzes has several benefits in addition to improving attendance and study behavior. These quizzes expose students to exam-type questions, help minimize test-taking anxiety, provide immediate feedback on students’ learning, identify areas of content difficulty to the instructor, and potentially improve long-term learning. An added benefit is that with infrequent administration, unannounced quizzes are simple to implement and require minimal time both in and out of the classroom.

References


Notes

1. Special thanks to David Hansen and Thea Rothmann for their outstanding editorial assistance.
2. Send correspondence to Haig Kouyoumdjian, Department of Psychology, University of Nebraska–Lincoln, Lincoln, NE 68588–0308; e-mail: hkouyoum@unlserve.unl.edu.

Differential Test Performance
From Differently Colored Paper:
White Paper Works Best

Nicholas F. Skinner
King’s College, The University of Western Ontario

Scores on multiple-choice tests printed on colored paper were uniformly lower than scores for tests printed on white paper. These results call into question the efficacy and fairness of the practice of attempting to reduce cheating by printing alternate forms of tests on paper of different colors.

One strategy for controlling cheating is printing tests on differently colored paper (e.g., Sinclair, Soldat, & Mark, 1998). However, studies of the effects of color on test performance have yielded contradictory results. Jacobs and Blandino’s (1992) findings using the Profile of Mood States (McNair, Lorr, & Droppleman, 1971) printed on white, blue, green, red, and yellow paper suggested the possibility of superior performance on tests printed on red and yellow paper because these colors decrease levels of test anxiety. On the other hand, Sinclair et al. (1998) reported better performance on blue paper than red paper in an actual examination setting. They explained this finding in terms of their cognitive tuning
This study examined whether offering daily extra credit quizzes predicted exam performance in an advanced psychology course (n = 36). Results revealed that extra credit performance was a strong predictor of exam performance, above and beyond gender, college grade point average, and ACT scores. In addition, results suggested that nearly half of the students reported their primary motivation for completing extra credit was to earn extra points toward their grade. This research has important pedagogical implications in highlighting the use of daily extra credit quizzes as a practice that facilitates learning and directly affects exam performance.

Teachers and researchers are interested in understanding what predicts academic success and in facilitating meaningful learning experiences for students. A number of student background variables are associated with academic achievement, such as ACT scores and college grade point average (GPA; Beecher & Fischer, 1999; DeBerard, Spielmans, & Julka, 2004). In addition, extra credit opportunities can be effective at encouraging class attendance (Wilder, Flood, & Stromsnes, 2001). However, researchers have not examined extra credit activities as direct predictors of academic achievement. Thus, this study examined the relation between daily extra credit quizzes and students’ exam performance.

The goal of extra credit in college courses is often to provide motivation for students to expand their understanding of course material and to enhance educational opportunities. The practice of offering extra credit is effective in motivating students to read journal articles (Carkenord, 1994), participate in classroom discussions (Boniecki & Moore, 2003), and attend class (Wilder et al., 2001). Another common way to motivate students to participate in class is through the use of required quizzes. Offering required quizzes during class improves student attendance (Hovell, Williams, & Semb, 1979), but students dislike the punitive approach of required quizzes (Combs, 1976), and similar outcomes can be obtained by instituting random extra credit quizzes (Thorne, 2000; Wilder et al., 2001). An option that has received less attention is offering daily extra credit quizzes designed to both encourage student attendance and increase student learning.

Critics of offering extra credit note that if only a select group of students participate in extra credit (even if it is available to all students), extra credit is not useful from a pedagogical standpoint, in that all students would not benefit from the learning experience (Hardy, 2002; Norcross, Dooley, & Stevenson, 1993). Indeed, research suggests that a small minority of students participate in common opportunities for extra credit, such as participating in research studies (Padilla-Walker, Zamboanga, Thompson, & Schmersal, 2005). However, the use of daily extra credit quizzes alleviates this concern, as all students who attend class participate, making this option desirable for instructors seeking to involve all of their students in an activity that may enhance learning.

Another related issue addressed in this study was students’ motivations for completing extra credit. Although Hill, Palladino, and Eison (1993) studied the forms of extra credit valued by instructors, what forms of extra credit students value is less clear. This question is important because the reasons for completion of extra credit assignments likely factor into the effectiveness of the assignments. Thus, in this study, students answered questions about their motivations for completing daily extra credit quizzes.

This study implemented the use of short, daily extra credit quizzes that focused on the readings for the current class period. Students responded to short-answer questions at the beginning of each class pertaining to the readings for that day. The primary purpose of this study was to determine the impact of offering daily extra credit quizzes on exam performance. Given that GPA and ACT are predictive of class performance, we expected that these background variables would predict extra credit performance. In turn, we hypothesized that performance on extra credit quizzes would predict exam scores, over and above college GPA and ACT scores. We also sought to understand if students’ motivation to complete quizzes was due to the oppor-
tunity to raise their course grade, the educational experience of participating in extra credit, or both.

Method

Participants

Participants were 36 undergraduate students enrolled in an advanced (400-level) child behavior and development course at a Midwestern state university (75% women; 90% third- and fourth-year students). This course was an upper level seminar course where students met twice a week for 75 min each time. Students completed daily textbook readings as well as daily readings that included empirical journal articles and book chapters. The course was a combination of lecture and discussion, but the focus was to involve students in the active discussion of theoretical and applied constructs relating to development.

Procedures

As instructor of the course, I used records of student exams and extra credit. After I obtained institutional review board approval and student consent, students completed a short questionnaire asking them to report their ACT score and college GPA as well as their motivations for completing daily extra credit quizzes. A research assistant administered the questionnaire on the last day of class, when I was not present.

Measures

Student background. Student background measures included self-reported ACT scores and college GPA.

Exam performance. I computed total exam scores as the sum of students’ scores on three short-answer and essay exams. Two were 50-point exams, plus a 75-point final exam.

Extra credit. At the beginning of the course I told the students that each day of class they would have an extra credit quiz on the readings for that day. Each quiz consisted of two short-answer questions (each worth 1/2 point of extra credit) based on the daily textbook or journal readings. Quizzes took approximately 5 min at the beginning of each class period. I graded each quiz question on a pass–no pass basis to simplify grading. I added points earned from extra credit quizzes to the final grades, with a total of 21 possible points of extra credit out of 350 total points for the class.

Student motivation for quizzes. Students ranked (from 1 to 5) the reasons why they valued the extra credit quizzes (e.g., “Because they helped me to prepare for the exam.”). I used these rankings to determine students’ primary motivations for completing extra credit quizzes.

Results

Descriptive Statistics and Tests of Gender Differences

I examined the data for univariate and multivariate outliers and identified no outliers. Table 1 contains the means, standard deviations, and ranges of the primary variables of interest. I conducted a number of univariate ANOVAs to determine if any of the variables varied as a function of gender. Of the four ANOVAs conducted, one was statistically significant. Women (M = 161.59, SD = 8.56) had higher overall exam scores than men (M = 149.11, SD = 10.78), F(1, 34) = 12.61, p < .001, partial η² = .27. An effect size of .27 is considered low to moderate (Cohen & Cohen, 1975), which suggests that the difference between exam scores for women and men may be meaningful. Thus, I included gender in subsequent analyses.

Variables Predicting Extra Credit Performance

I conducted a multiple regression analysis to determine which variables predicted extra credit performance. Predictors entered simultaneously were gender, college GPA, and ACT scores. The regression predicting extra credit performance accounted for a statistically significant amount of total variance, R² = .35, F(3, 32) = 5.76, p < .01, with college GPA positively predicting extra credit performance (β = .61, p < .001). Gender (β = .20) and ACT scores (β = -.29) did not significantly predict extra credit performance.

Variables Predicting Exam Performance

I conducted a multiple regression analysis to determine whether extra credit performance predicted overall exam performance above and beyond gender of the student, college GPA, and ACT scores. Predictors entered simultaneously were gender, college GPA, ACT scores, and extra credit performance. The regression predicting exam scores accounted for a statistically significant amount of total variance, R² = .47.

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>College GPA</td>
<td>3.27</td>
<td>0.32</td>
<td>2.60 to 3.99</td>
</tr>
<tr>
<td>ACT score</td>
<td>24.03</td>
<td>3.58</td>
<td>18 to 31</td>
</tr>
<tr>
<td>Total exam grade</td>
<td>158.47</td>
<td>10.54</td>
<td>138 to 173</td>
</tr>
<tr>
<td>Total extra credit</td>
<td>12.71</td>
<td>4.33</td>
<td>4.50 to 20.75</td>
</tr>
</tbody>
</table>

Note. GPA = grade point average.
Students' primary motivations for completing extra credit quizzes. Almost half (49%) of the students reported valuing the extra credit quizzes primarily as a means of raising their course grade. About one fifth of students (22%) reported valuing extra credit quizzes as motivation to read course material, 17% as motivation to come to class and keep up, 11% because the quizzes were helpful in learning the course material, and 3% because the quizzes were helpful in preparing for the exam.

Discussion

The primary purpose of this study was to determine if performance on daily extra credit quizzes predicted exam performance. In addition, this study examined students' motivations for completing extra credit quizzes. Findings suggested that extra credit performance predicted exam performance over and above gender of the student, college GPA, and ACT scores. In addition, nearly half of the students reported that their primary motivation for completing extra credit was to earn points toward their course grade.

Variables Predicting Extra Credit and Exam Performance

Adding to research suggesting that prior GPA is associated with academic achievement (DeBerard et al., 2004), this study found that college GPA was a strong predictor of performance on extra credit quizzes. It is possible that students who were motivated to do well in class were also motivated to keep up in class, which facilitated performance on extra credit quizzes. That being said, performance on extra credit quizzes was a direct and strong predictor of exam performance, even after controlling for gender, college GPA, and ACT scores. These findings suggest that the form of extra credit used in this study was indeed effective at facilitating mastery of course material, beyond the variance accounted for by background variables. However, completing the class readings that were necessary to do well on the quizzes likely facilitated better understanding of the material and stronger exam performance. Future research should systematically consider the role extra credit quizzes may have on students' study habits and how study habits impact exam performance.

Student Motivations for Completing Extra Credit Quizzes

Although some research suggests that students' reasons for completing extra credit may be to boost or secure their course grades (Bender, 1986; Padilla-Walker et al., 2005), little research has directly addressed this topic. Although almost half (49%) of students reported being motivated to complete extra credit quizzes for the extra points toward their grade, many students also reported that the quizzes were effective at encouraging them to read course materials and attend class. Unlike other forms of extra credit, students valued extra credit quizzes as a tool motivating them to read class materials and stay caught up in a challenging course with heavy reading assignments. Thus, this form of extra credit is attractive for students who are seeking to secure their course grade. In addition, daily extra credit quizzes are attractive pedagogically because course grades are not inflated (the average extra credit points earned over the entire semester was 12.71) and all students in the course participate to a large degree. These characteristics make this option a viable form of extra credit, even for those who typically criticize the use of extra credit.

Conclusions

Future research is needed comparing courses that use this procedure to those that do not to rule out the possibility that this effect was unique to the current sample, although that possibility is unlikely. Indeed, I have used this extra credit procedure with freshman students as well as with larger classes and have found similar results supporting the effectiveness of this practice at facilitating learning and understanding of course topics. When using this technique in larger classes, I use multiple-choice questions to simplify grading, although it is possible that to simplify grading further, random extra credit quizzes would be as effective as daily quizzes. This possibility warrants future research with larger samples.

Overall, this study provided support for the pedagogical utility of offering daily extra credit quizzes in an advanced level psychology course. Although not overinflating students' grades, students who did well on extra credit quizzes also performed better on exams, above and beyond general measures of achievement. In summary, students perceive positive incentives (as opposed to punishing them with pop quizzes) in a positive manner. These positive incentives encourage students to read their daily assignments, which improves their overall understanding of the material and facilitates learning.
References


Notes

1. I gathered these data while at the Department of Psychology, University of Nebraska–Lincoln.

2. I greatly appreciated the helpful feedback of Byron Zamboanga on earlier versions of this article.

3. Send correspondence concerning this article to Laura M. Padilla-Walker, School of Family Life, 2097 JFSB, Brigham Young University, Provo, UT 84602; e-mail: laura_walker@byu.edu.
Introductory Psychology Student Performance: Weekly Quizzes Followed by a Cumulative Final Exam

R. Eric Landrum
Boise State University

Students in an introductory psychology course took a quiz a week over each textbook chapter, followed by a cumulative final exam. Students missing a quiz in class could make up a quiz at any time during the semester, and answers to quiz items were available to students prior to the cumulative final exam. The cumulative final exam consisted of half the items previously presented on quizzes; half of those items had the response options scrambled. The performance on similar items on the cumulative final was slightly higher than on the original quiz, and scrambling the response options had little effect. Students strongly supported the quiz a week approach.

The introductory psychology course is a popular course for nonpsychology majors completing general education (or core) requirements and a required course for nearly all undergraduate psychology majors nationwide (Perlman & McCann, 1999). In my previous approach to teaching the course, I divided the content into “units” covering two to four chapters each, followed by a major exam, with the course containing five or six units and exams. This format led to large reading assignments for students (often more than 100 pages). Given that researchers have found that introductory students retain little after the semester is over (Rickard, Rogers, Ellis, & Beidleman, 1988; VanderStoep, Fagerlin, & Feenstra, 2000), I began thinking about alternative approaches to teaching the course. I wanted to redesign the course to reduce student anxiety and enhance student performance, so I changed my teaching and testing strategy dramatically.

I implemented a teaching approach that covered one chapter per week with a quiz at the end of each week. At the end of the course, I gave students a cumulative final exam. In an effort to reduce anxiety, I also implemented a lenient and generous make-up policy, such that I allowed students to make up quizzes at any time during the semester prior to the last day of class. Students knew from the beginning of the course that from each 20-item weekly quiz, 10 items would appear again on the cumulative final exam; students could also retrieve (from teaching assistants) a printout with their quiz answers and the correct answers. Of the items appearing on the cumulative final exam, half of the items (75) had scrambled response options (a, b, c) compared to the original quiz item. Although software programs are certainly capable of scrambling response options (e.g., Harnisch & Rotheroe, 1986), to my knowledge there is no information available about the change in item correctness (proportion of students answering an item correctly) when response options are scrambled in a test–retest situation.

Some of the inspiration for making the course changes came from Weimer (2002) and her work on learner-centered teaching; other researchers have tested different ideas. For instance, Grover, Becker, and Davis (1989) found that students preferred a frequent testing program (chapter by chapter) over a unit testing program (four chapters). Although student performance was similar for both approaches, only students in the frequent testing program indicated that they would choose that option again. Grabe (1994) found that whether students had one or more attempts at taking exams earlier in the course had no effect on cumulative final exam performance. Kahn (2000) found that students who missed required exams scored lower on comprehensive final tests and had more overall class absences.
I was interested in answering the following questions: (a) What is the relation between a lenient (i.e., generous) make-up quiz policy and student scores? (b) How was overall quiz performance related to cumulative final exam performance? (c) What is the effect of scrambling the response options for half of the cumulative final exam items?

**Method**

**Participants**

General psychology students (N = 253) enrolled in my Fall 2003 course were the participants in this study. Of the 3,795 possible quizzes (15 quizzes × 253 students), students completed 3,385 quizzes during the scheduled quiz time (89.2%). Students completed 249 quizzes as make-up quizzes (7.4%); 161 missed quizzes were not made up (4.2%).

**Materials**

I created a 20-item multiple-choice quiz with three response options (a, b, and c) for each chapter every week of the course. Each week covered one textbook chapter from Lahey (2004). Students completed quizzes on Scantron bubble sheets; the following week students could pick up printouts of their quiz performance indicating their answers and the correct answers. At the end of the semester students completed a cumulative final exam. The items for the cumulative final exam appeared in the same sequence as the topics appeared during the course.

**Procedure**

Students completed weekly quizzes in class. I told students at the beginning of the semester that the quiz items would comprise the cumulative final exam items, but that some of the response options would be scrambled.

**Results and Discussion**

First, I examined make-up quiz performance versus quiz performance of students taking the weekly quiz at the scheduled time. Because make-up quizzes were hand-scored, these quizzes were easily identifiable. Second, I compared average quiz performance scores with performance on the cumulative final exam.

Lastly, I analyzed item correctness statistics of the original quiz items compared to the repeated cumulative exam items, attempting to detect the effect (if any) of response option scrambling.

**Make-Up Quiz Performance Versus Regularly Scheduled Quiz Performance**

I conducted a 15 (weekly quiz) × 2 (quiz type: make-up vs. in class) ANOVA to examine quiz performance. There was a significant main effect of quiz, $F(14, 3355) = 8.40$, $p < .001$. Quiz 5 performance was unusually low, and Quizzes 1, 2, 4, and 7 exhibited high overall scores. There was a significant main effect of quiz type, $F(1, 3355) = 6.01$, $p < .02$. Overall, students taking make-up quizzes scored significantly higher ($M = 15.3, SD = 3.4$) than students taking the quiz at the regularly scheduled time ($M = 14.7, SD = 3.0$).

There was also a significant interaction between the weekly quiz and quiz type, $F(14, 3355) = 2.58$, $p < .005$ (see Table 1). The pattern of change over time did not appear to be stable. In 9 of 15 quizzes, students in the make-up condition performed better; in the remaining 6 of 15 quizzes, students taking the quiz in class performed better. This unpredictable pattern of quiz performance may be due to the varying difficulty of chapter topics covered each week.

**Average Quiz Performance Compared to Cumulative Final Exam Performance**

I calculated an average quiz score for each student and compared it with each student’s cumulative final exam score. There was a significant correlation between average quiz score and cumulative final exam score, $r(222) = 0.65$, $p < .001$. I then divided students, according to their quiz score averages, into thirds (top, middle, bottom) and then compared their quiz score ranks with cumulative final exam performance to test the notion that “those who have the most to gain, gain the most.” After converting average quiz score and cumulative final exam scores into percentages, I performed a 3 (top, middle, bottom) × 2 (percent quiz score, percent final exam score) repeated measures ANOVA to detect differential changes across test performance dependent on the ranks. As expected, the rank-ordered groups did differ on final exam performance (top third, $M = 95.4, SD = 5.5$; middle third, $M = 86.8, SD = 7.7$; bottom third, $M = 78.6, SD = 12.7$), $F(2, 214) = 220.58$, $p < .001$. For all groups, there was an overall increase in percentage correct from
Table 1. Means and Standard Deviations for Quiz Scores by Week and Quiz Type

<table>
<thead>
<tr>
<th>Quiz Week</th>
<th>In class</th>
<th>Make UP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1</td>
<td>16.07</td>
<td>2.61</td>
</tr>
<tr>
<td>2</td>
<td>17.15</td>
<td>2.59</td>
</tr>
<tr>
<td>3</td>
<td>14.48</td>
<td>2.65</td>
</tr>
<tr>
<td>4</td>
<td>15.38</td>
<td>2.66</td>
</tr>
<tr>
<td>5</td>
<td>12.97</td>
<td>2.93</td>
</tr>
<tr>
<td>6</td>
<td>13.36</td>
<td>3.44</td>
</tr>
<tr>
<td>7</td>
<td>16.98</td>
<td>2.49</td>
</tr>
<tr>
<td>8</td>
<td>14.39</td>
<td>2.75</td>
</tr>
<tr>
<td>9</td>
<td>13.89</td>
<td>3.07</td>
</tr>
<tr>
<td>10</td>
<td>15.42</td>
<td>2.61</td>
</tr>
<tr>
<td>11</td>
<td>13.13</td>
<td>2.93</td>
</tr>
<tr>
<td>12</td>
<td>15.87</td>
<td>1.95</td>
</tr>
<tr>
<td>13</td>
<td>14.03</td>
<td>2.88</td>
</tr>
<tr>
<td>14</td>
<td>14.18</td>
<td>3.01</td>
</tr>
<tr>
<td>15</td>
<td>14.56</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Did response option scrambling have any effect on changes in item correctness? I compared performance on the cumulative final exam to performance on the original quiz and calculated a change score for each of the 150 items. A positive change score indicated that the cumulative final exam item was easier to answer than the original quiz item. Items that kept the identical response option order from original quiz to cumulative final exam changed in item correctness an average of 0.067 (SD = 0.08), indicating that the cumulative final exam items were easier, answered on average 6.7% more accurately than previously. Items that had their response options scrambled from the original quiz to the cumulative final exam changed in item correctness an average of 0.060 (SD = 0.81), meaning that the scrambled items were also easier on the cumulative final exam, answered on average 6.0% more accurately than previously. Although the growth in increased performance for scrambled items was not as high for those with intact response options, this difference was not significant, t(149) = 0.47, ns. Scrambling the response options did not make cumulative final exam questions significantly harder; performance improved for both types of items (original and scrambled).

Conclusions

I was surprised that scores on the make-up quizzes were significantly higher than those on the in-class quizzes. Given the rush to take make-up quizzes at the quizzes (M = 74.0, SD = 9.8) to the cumulative final exam (M = 86.9, SD = 11.4). However, there was also a significant interaction between quiz rank and change from quiz percentage correct to final exam percentage correct, F(2, 214) = 6.89, p < .002. The lowest third of quiz scorers increased from an average of 62.8% correct (SD = 4.8) to an average of 78.6% correct on the final exam (SD = 12.7). Middle-ranked quiz scorers increased from 74.2% quiz performance (SD = 2.3) to 86.8% final exam performance (SD = 7.7), and top-ranked quiz scorers increased from 85.0% correct on quiz performance (SD = 4.0) to 95.4% correct on final exam performance (SD = 5.5). Tukey’s post-hoc tests indicated that all of the quiz–final exam paired means were significantly different from one another, p < .05.

Effects of Response Option Scrambling on Changes in Item Correctness

The 150 items on the cumulative final exam had an initial average item correctness of .81 (SD = 0.13). That is, students answered these items correctly 81.0% of the time when they appeared on weekly quizzes. When these items appeared again on the cumulative final exam (with scrambled response options), item correctness increased to .87 (SD = 0.09). This change was significant, t(149) = –9.49, p < .001. Thus, students answered the reoccurring multiple-choice items more accurately on the cumulative final exam (average 87.4% correct) than on the original 15 quizzes (average 81.0% correct).
end of the semester, I expected that students did not prepare well and crammed these quizzes in at the last minute. Perhaps the additional time allowed students to prepare better; I also must consider that because previous quiz answers were available, students shared quiz results. In the future, I will use a different make-up strategy.

No matter what a student's quiz percentage ranking (top third, middle third, bottom third), student scores increased when comparing the change from overall quiz percentage to overall final exam percentage. However, the significant interaction indicated that those students with the most to gain (the lowest third) did gain the most. When comparing these percentages, students in the top third experienced a 10.4% growth in scores; the middle third, 12.6% growth; and the bottom third, 15.8% growth. The growth in the top third, however, may be limited in part to a ceiling effect (this group scored 95.4% correct on their cumulative final exam).

When comparing the same item correctness from original quizzes to cumulative final, students performed significantly better on the cumulative final exam (87.4% correct) compared to the original quiz items (81.0% correct). Given the changes that I made to my introductory course (e.g., one chapter and one quiz every week), I expected that students might not like this teaching approach. This particular course met Monday, Wednesday, and Friday at 8:40 a.m., and I gave a quiz every Friday at the end of class. About midway through the semester on a nonquiz day, I gave students attending class that day (n = 165) a quiz every Friday at the end of class. About midway through the semester on a nonquiz day, I asked students attending class that day (n = 165) if I should continue using the quiz every week approach. Of those in attendance, 92.7% replied yes, 5.5% replied no, and 1.8% replied with a “don’t know” or were nonresponsive.

I encourage other faculty members to experiment with their classes and try new approaches. I thought I was taking a substantial risk in redesigning my course, and I thought that the change might hinder student performance and student opinion might be negative. The results of this study indicate that I still have as much to learn from my students as they do from me.

References


Notes

1. I appreciate earlier comments from the Editor and the three reviewers.

2. Send correspondence to R. Eric Landrum, Department of Psychology, Boise State University, 1910 University Drive, Boise, ID 83725-1715; e-mail: elandru@boisestate.edu.
I tested whether students' reported use and perceived helpfulness of textbook pedagogical aids and other instructional methods (e.g., group discussion) related to better exam performance. More than 200 undergraduates rated 10 commonly used pedagogical aids and instructional techniques. Students reported using boldface and italicized terms and practice test questions most often, and they considered boldface and italicized terms, practice test questions, and online quizzes most helpful. The reported use of pedagogical aids and their perceived helpfulness did not relate to student performance on exams. Furthermore, the rated helpfulness of key terms related to worse scores on exams. Pedagogical aids may be useful teaching tools and learning enhancers, but this commonsensical hypothesis must be tested.

Introductory psychology textbooks feature many pedagogical aids to enhance student learning. Ranging from boldface terms to chapter summaries and key term sections, these aids have grown in number and variety over the decades (Marek, Griggs, & Christopher, 1999; Weiten & Wight, 1992). These learning aids are designed to help students master the material and provide an additional variable for instructors to compare and contrast the multitude of textbooks on the market (e.g., Griggs, 1999; Griggs, Jackson, & Napolitano, 1994). In fact, Weiten (1988) showed that the use of italics, a common pedagogical aid, predicted instructors' beliefs in a textbook's capacity to engage student interest and their awareness of a book, although the aids did not relate to professors' evaluations of a text's overall pedagogical quality.

The potential importance of pedagogical aids has compelled a number of empirical studies of students' perceptions of these aids (Marek et al., 1999; Weiten, Deguara, Rehmke, & Sewell, 1999; Weiten, Guadagno, & Beck, 1996), but little is known about their effectiveness. For example, Weiten et al. (1996) showed that the aids varied considerably in terms of their familiarity to students, their likelihood of being used, and their perceived value. They also found that students' educational experience, as measured by their year in school and their previous academic grade point averages, did not relate strongly to their ratings of pedagogical aids. Similarly, Marek et al. showed that some aids are both more prevalent and considered most valuable (e.g., boldface) and that first-semester and advanced students tended to rate pedagogical aids in a similar fashion. Weiten et al. extended these findings, showing that university, community college, and high school students agreed about which learning aids were most useful. Knowing about students' preferences is useful in predicting the extent to which they may use particular aids and which aids may be helpful, but how much do students actually use these aids? Furthermore, does their use correlate with standard measures of learning (i.e., exam scores)? This study extends previous findings by assessing how the use of different aids and their rated helpfulness vary and relate to exam performance.

Textbook pedagogical aids are particularly important in large introductory classes where the standard lecture is often the most common and most practical way of interacting with students. Many instructors do not have the time or training to conduct in-class exercises to engage different levels of understanding and different types of learning styles in the way that many pedagogical aids can. Having effective pedagogical aids can benefit both the instructor and the student, and aids should thereby be a valid criterion to compare textbooks. Conversely, it is likely that some pedagogical aids serve as mere window dressing, conveying the impression that they enhance a textbook to better serve students, but serve no real function. Furthermore, the use of some aids might actually hurt students who take time to use them instead of spending their time in a more productive way. This study correlated the use of different textbook aids with students' scores on exams to ascertain the effectiveness of each aid. It also compared the use of textbook aids with other instructional techniques such as group exercises, a paper assignment, and online quizzes. This information can help instructors in their choice of textbooks, their choice and implementation of pedagogical aids, and in guiding their students in how to best study.

Method

Participants

Two hundred and thirty-seven students (179 women and 58 men) from a midsized midwestern university in two sections of my introductory psychology class participated in this study (participation was voluntary). The majority of the students were first-semester freshmen (90%). The remaining were second-year students (6%), third-year students (3%), and seniors (1%). I combined the data from both classes as exam grades and results did not differ between sections.
Survey Materials

The survey measured the extent of use and the perceived effectiveness of six pedagogical aids present in the assigned textbook (chapter outlines, chapter summary/review sections, boldface terms, italicized terms, key terms, and practice test questions). Each aid was defined based on a pedagogical aids survey originally administered by Weiten et al. (1996). Participants first read a brief description of the aid and then used a 4-point scale to indicate the extent to which they used the aid with scale points at 1 (not at all/do not remember seeing it), 2 (a little/looked at it sometimes), 3 (a moderate amount/used it often), and 4 (extremely/used it all the time). Participants then rated the overall value of each aid on a 4-point scale ranging from 1 (not helpful) to 4 (extremely helpful).

The textbook assigned to the class (Weiten, 2002) was a brief version of a full-length textbook in the high middle level of difficulty (Griggs, 1999) and had a large number of common pedagogical aids. Two undergraduate teaching assistants and I compared this text with other textbooks at the same difficulty level (e.g., Griggs, 1999) and found enough similarities with them to suggest that findings from this study should generalize to pedagogical aids in other books.

I also assessed other instructional methods that I built into the class structure. Participants rated their use of online quizzes, class group exercises, research participation, and a class paper assignment. To ensure that students were familiar with the material before coming to class, I required students to complete an online quiz (from the text Web site or written by me) on each chapter in the textbook before I discussed it in class. I also conducted in-class group exercises in which groups of 5 students completed applied exercises relating to the material discussed in class the previous week. I also required students to take part in one research study during the semester and asked them to write a three-page paper using as many different areas of psychology as possible to explain an issue or phenomenon of their choice (e.g., violence in schools).

Procedure

I added the survey to the last of four exams, which occurred during finals week. After 65 multiple-choice questions testing material from class, participants read instructions stating that the remaining questions on the exam sheet would assess how the various parts of the class and the textbook influenced learning. I told students that participation was voluntary and that the answers to the questions would not affect their class grades or exam scores. Instructions also stated that I would discuss only group-level information from the questions and their responses would remain confidential. All students consented to participate.

Results and Discussion

Tables 1 and 2 present the average reported use and average reported utility data for each textbook pedagogical aid and the percentage of students who chose each response option for the use and utility of each aid. Students reported using boldface and italicized terms most often, followed by the practice test questions, the chapter summaries/reviews, the key terms, and the chapter outlines. A series of paired sample t tests showed that students reported using aids to different extents, except for the use of chapter summaries/reviews and the use of practice test questions (both used to the same extent). Many authors (e.g., Marek et al., 1999) have suggested that students pay more attention to those aids that directly relate to exams, and this claim was mostly true in this sample (60% of the class used the practice test questions moderately or extremely).

The most helpful pedagogical aids were boldface terms, italicized terms, and practice test questions (92% of the sample rated boldface terms as moderately or extremely helpful). Again, students rated most aids significantly differently in helpfulness from each other. Only the reported helpfulness of the chapter summary/review sections and italicized text aids, and the italicized and practice test question aids, did not differ significantly from each other.

Note. Online quizzes, group work, research, and papers were all required (i.e., no use measure). Ratings based on a scale ranging from 1 (not at all/do not remember seeing it), 2 (little), 3 (moderate), to 4 (extremely/used it all the time). These represent average student ratings (N = 237). The reported use of aids sharing a subscript was not significantly different from each other.

Table 1. Descriptive Data for Reported Use of the Different Pedagogical Aids

<table>
<thead>
<tr>
<th>Frequency of Use (%)</th>
<th>M 1 2 3 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlines</td>
<td>1.93</td>
</tr>
<tr>
<td>Chapter summary/reviews</td>
<td>2.69</td>
</tr>
<tr>
<td>Boldface terms</td>
<td>3.52</td>
</tr>
<tr>
<td>Italics terms</td>
<td>3.12</td>
</tr>
<tr>
<td>Key terms</td>
<td>2.22</td>
</tr>
<tr>
<td>Practice test questions</td>
<td>2.86</td>
</tr>
</tbody>
</table>

Table 2. Descriptive Data for Helpfulness of the Different Pedagogical Aids

<table>
<thead>
<tr>
<th>Frequency for Helpfulness Ratings (%)</th>
<th>M 1 2 3 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlines</td>
<td>2.30</td>
</tr>
<tr>
<td>Chapter summary/reviews</td>
<td>2.99</td>
</tr>
<tr>
<td>Boldface terms</td>
<td>3.58</td>
</tr>
<tr>
<td>Italics terms</td>
<td>3.18</td>
</tr>
<tr>
<td>Key terms</td>
<td>2.65</td>
</tr>
<tr>
<td>Practice test questions</td>
<td>3.18</td>
</tr>
<tr>
<td>Online quizzes</td>
<td>3.13</td>
</tr>
<tr>
<td>Group exercises</td>
<td>2.76</td>
</tr>
<tr>
<td>Research participation</td>
<td>2.41</td>
</tr>
<tr>
<td>Paper assignments</td>
<td>2.92</td>
</tr>
</tbody>
</table>

Note. Helpfulness ratings based on a scale ranging from 1 (not at all), 2 (somewhat), 3 (moderate), to 4 (extremely helpful). These represent average student ratings (N = 237). The reported helpfulness of aids sharing a subscript was not significantly different from each other.
Students rated four of the textbook pedagogical aids (all excluding the chapter outlines and key terms) as more helpful to their learning than the in-class group exercises, the class paper assignment, or the research participation assignment. These instructional methods required more time and effort on the part of both the instructor and students and are elaborate forms of active learning. Although the lower helpfulness ratings of these techniques do not warrant their elimination, instructors need to find ways to enhance the helpfulness of such exercises and to assess their effectiveness.

The results of zero-order correlations between the reported use of different aids showed that the students’ reported use of one aid did not guarantee the reported use of other aids (data available from the author on request). Most correlations within usage categories were low to moderate. The reported use of chapter summary/review sections correlated with reported use of all other aids except for italics. Students who reported using the chapter summary/review sections were particularly likely to report using the practice test questions, \( r(237) = .53, p < .01 \). Similarly, students who reported using boldface terms were quite likely to report using italicized terms, \( r(237) = .54, p < .01 \). Correlations between the ratings of helpfulness of different aids showed similar results. Most correlations within helpfulness categories were low with some exceptions. Students who found the chapter summary/review sections helpful were particularly likely to find the practice test questions helpful, \( r(237) = .52, p < .01 \). Similarly, students finding boldface terms helpful were extremely likely to find italicized terms helpful, \( r(237) = .62, p < .01 \).

The correlations between reported usage of aids and their helpfulness were moderate to strong. For example, as can be expected, students reporting use of boldface terms also rated them as being very helpful, \( r(237) = .66, p < .01 \). I found only limited associations between the reported use of textbook pedagogical aids and the helpfulness of instructional methods. Students who rated doing research as being a helpful learning experience also were more likely to report using outlines, \( r(237) = .31, p < .01 \); boldface terms, \( r(237) = .20, p < .01 \); and key terms, \( r(237) = .25, p < .01 \).

Did the reported use of pedagogical aids predict exam scores? Zero-order correlations assessed whether the reported use of different aids and their perceived helpfulness related to performance on exams in class. I averaged the four exams to create a reliable overall index of student test performance (Cronbach’s \( \alpha = .92 \)) and reduce the likelihood of observing spurious significance in a large set of correlations. In general, the reported use and the perceived helpfulness of the different pedagogical aids were not significantly associated with exam scores. Only one correlation was significant. Student ratings of the helpfulness of key terms was negatively related to their exam scores, \( r(237) = -.20, p < .01 \). A closer look at this finding with an ANOVA showed that students who reported the key terms as being more helpful had significantly lower exam scores, \( F(3, 228) = 4.35, p < .01 \). Students who did not find key terms at all helpful achieved a mean of 85 points compared to students who found key terms most helpful, who scored 77 points on average, \( t(225) = 2.63, p < .01 \).

Certain limitations of this study should temper the impulse to ignore pedagogical aids. I did not control for two factors that could account for a large amount of variance in exam performance, student ability and student effort or time studied. Additionally, the analyses involved a large number of correlations and many of the associations were low to moderate in magnitude. These findings may vary for texts of different difficulty levels. Finally, the retrospective and cross-sectional nature of the data precludes drawing causal conclusions. The measure of “use” could be particularly inaccurate, as use could have varied during the 16-week semester and could be biased by student impression management or demand characteristics of the setting. Prospective studies of students’ reported use of pedagogical aids combined with measurements of actual use will greatly enhance the assessment of pedagogical aids and techniques. The results nonetheless provide a compelling set of information urging more research on this topic.

Conclusions

The number of pedagogical aids in a textbook influences how a book looks, in terms of visual appearance, size, and perceived ease of use. Instructors expect pedagogical aids to increase the elaboration of material, test understanding, and enhance learning, but this effect was not seen in the most common measure of learning, exam performance. Expanding on extant knowledge of students’ perceptions of pedagogical aids, this study showed that students reported using aids such as practice test questions and text fonts and considered them helpful to their learning, but simply using an aid did not strongly relate to the students considering the aid helpful or to students’ exam scores.

Should adoption of a certain textbook depend on whether it has the well-used aids? Apart from the fact that almost all textbooks have aids like boldface and italics (Griggs et al., 1994), correlational analyses did not show any positive relations between the reported use of pedagogical aids and learning as measured by exam performance. The results suggest that absence of the aids surveyed should not be grounds for avoiding a text. Additionally, textbook authors should not feel pressured to load their books with such aids. Although conveying the appearance that the book has many additional ways to facilitate learning, high- and low-scoring students did not use the pedagogical aids differentially, nor did the use of aids increase exam performance.

The lack of effectiveness of textbook pedagogical aids on exam performance is not an isolated finding. A recent study by Balch (2001) compared the utility of different study tips (e.g., course performance charts, study groups, lecture notes, self-help quizzes) and found a similar low relation between aids and course performance. Results such as these call for instructors to pay closer attention to the different pedagogical aids on which they rely or urge students to use. Many books help students pick the best ways to study (e.g., Heffernan, 2000; Thomas & Rohwer, 1986); this study compels a closer look at the empirical evidence supporting book recommendations.

These findings also urge instructors to look at how learning and teaching are designed and assessed. As suggested in the backward design method of instruction and class planning, teaching techniques should be guided by first setting clear desired goals and standards, then designing accurate perfor-
performance measures of the goals, and finally planning learning experiences and instruction to match (Wiggins & McTighe, 1998). Multiple-choice exams are one of the most common assessment methods used in large introductory psychology classes, but they assess only limited forms of learning. Most multiple-choice exams are not written to assess students’ ability to analyze, synthesize, evaluate, or apply their knowledge (Appleby, 2001). Although my exam questions tested whether students could apply the knowledge they had in addition to recalling facts, they may not have assessed other forms of learning enhanced by the different pedagogical aids. It is possible that different assessment measures could reveal that pedagogical aids do enhance learning. A better assessment of the effectiveness of aids and other instructional techniques could be achieved by an examination of different outcome measures (e.g., content of group exercise discussions, quality of critical thinking demonstrated in the papers).

Can using some pedagogical aids actually hurt exam performance? The correlational nature of the data does not allow for a true test of this question, and the single significant finding (relating to a less-used aid) does not warrant alarm. Nonetheless, the misuse or overuse of pedagogical aids can be potentially damaging. Students could spend too much time on some aids (e.g., key terms) at the expense of studying other important material or working on elaboration and understanding of the material. For example, Brothen and Wambach (2001) found that spending more time taking quizzes related to poorer exam performance. These results, together with the knowledge that students are selective in what aids they use and find helpful, has ramifications for how instructors advise students on their study habits. Instructors should provide explicit guidelines on how to use pedagogical aids and work to enhance the aids’ helpfulness and effectiveness. Instructors and textbook study tips sections should caution students against the tendency to just focus on one type of aid and should urge students to use aids that they may not be using enough (e.g., practice test questions).

References


Notes

1. The writing of this article was partially funded by the University of Wisconsin System Teaching Fellows Program and the Office of Professional and Instructional Development (OPID).

2. I thank Martha Ahrendt, Randolph Smith, and anonymous reviewers for their comments and suggestions; Jeffrey Zellner and Beckie Zingler for their help with collecting the data; and Twila Marquardt for her help with data management.

3. Send correspondence to Regan A. R. Gurung, Department of Human Development and Psychology, University of Wisconsin, Green Bay, 2420 Nicolet Drive, MAC C318, Green Bay, WI 54311; e-mail: gurungr@uwgb.edu.
Pedagogical Aids: Learning Enhancers or Dangerous Detours?

Regan A. R. Gurung
Human Development and Psychology
University of Wisconsin, Green Bay

I tested whether students’ scores on the semester’s first exam would change their reported use of textbook pedagogical aids during the course of the semester. I assessed 240 introductory psychology undergraduates early and late in the semester. The reported use of aids did not positively relate to student performance on any exams and showed only limited changes over the course of the semester. Poor scores on the first exam led only to small increases in the use of summary sections. Furthermore, the use of aids such as boldface terms decreased. Men and more able students used aids less, and women showed greater decreases in the perceived helpfulness of some aids over the semester.

Instructors provide a variety of learning aids to students. Some students rely heavily on them, and others rarely take full advantage of them. The overuse of some aids could keep students from studying in more effective ways. Conversely, some students do not use learning aids enough because they believe that they do not need them or that the aids will not be helpful. Students often believe that reading the textbook and their notes is sufficient preparation for an exam (Cross & Steadman, 1996). If proven wrong (i.e., they score badly on an exam), would students increase their use of pedagogical aids? This article addresses this question and highlights pedagogical aids’ role as a learning enhancer.

Pedagogical aids are designed to be learning enhancers. Optimally, aids serve as metacognitive tools alerting students as to how much they know and how well they know it as well as giving them the ability to consciously and deliberately monitor and regulate their learning (Hacker, Dunlosky, & Graesser, 1998). Every introductory psychology book includes a variety of study aids designed to help students enhance their learning. Pedagogical aids have populated texts for decades (Marek, Griggs, & Christopher, 1999), and a number of empirical studies have assessed students’ perceptions of these aids (e.g., Gurung, 2003; Weiten, Deguara, Rehmke, & Sewell, 1999). Although it is clear that students use many of the aids in texts (Weiten, Guadagni, & Beck, 1996), faculty lack empirical data on the effectiveness of pedagogical aids. One study showed that use does not seem to relate to better exam performance, and the perceived helpfulness of aids such as key terms was associated with worse exam performance (Gurung, 2003).

The potential importance of pedagogical aids compels a thorough understanding of their effectiveness and use. Pedagogical aids, if used correctly, could alert students to what they know or do not know and enhance their learning. Alternatively, the use of some aids might actually hurt students who use them incorrectly and waste study time memorizing the aids instead of using the aids as a test of knowledge (i.e., a dangerous detour). Unfortunately, many students may use aids only because they get credit for doing so or may memorize the information in an aid perceiving those specific items likely to be on exams (Marek et al., 1999). The first exam is the time that most students get a true sense of what they know and whether their studying methods were sufficient. Correspondingly, I hypothesized that students who scored poorly on an early exam would use aids more for the rest of the semester. No extant work has assessed how aid usage varies over a semester. I used a prospective design and controlled for ability and effort, unlike previous cross-sectional research (e.g., Gurung, 2003), to assess change in aid use and to expand on knowledge of who uses pedagogical aids. I also tested for factors, such as ability, that predicted change in aid usage.

Method

Participants

Two hundred and forty students (164 women and 76 men) from a midsized midwestern university in two sections of my introductory psychology class participated in this study (participation was voluntary). Twenty-four students did not complete both measurements and were excluded from the study. The mean age was 20.08 (SD = 4.20). The majority of the students were freshmen (63%); the remainder were sophomores (17%), juniors (10%), and seniors (10%). I combined the data from both sections as exam grades were similar.

Materials

A short survey measured the extent of use and the perceived effectiveness of five pedagogical aids present in the assigned textbook (chapter summary sections, boldface terms, italicized terms, key terms at the end of chapters, practice test questions). Participants first read a brief description of the aid (see Weiten et al., 1996) and then used a 4-point scale to indi-
cated the extent to which they used the aid: 1 (not at all/not remember seeing it), 2 (a little/looked at it sometimes), 3 (a moderate amount/used it often), 4 (extremely/used it all the time). Participants then rated the overall value of each aid on a 4-point scale ranging from 1 (not helpful) to 4 (extremely helpful). I also assessed other instructional methods that I built into the class structure (i.e., those not tested in Weiten et al., 1996). Participants rated their use of online quizzes (textbook publisher’s site), class group exercises, research participation, and a class paper assignment. I asked students the number of hours they studied for the exam to measure effort, and I used ACT scores and high school GPA to control for ability.

Procedure

I added the survey to the end of the first and last of four exams. After answering 65 multiple-choice questions, participants read instructions stating that the remaining questions on the exam sheet would assess how the various parts of the class and the textbook influenced learning. I told students that participation was voluntary and that the answers to the questions would not affect their class grades or exam scores.

Results and Discussion

The average reported use and perceived helpfulness for each textbook pedagogical aid for each assessment period was consistent with previous research (i.e., Gurung, 2003; tables available on request). Students reported using boldface and italicized terms most often, followed by the practice test questions. There was a significant correlation between number of hours studied (i.e., effort) and Exam 1 scores, \( r(216) = .15, p < .05 \).

Do First Exam Scores Predict Later Studying?

Partial correlations controlling for effort and ability (high school GPA and ACT scores) assessed whether performance on the first exam related to the reported use of different aids and their perceived helpfulness at the second assessment and to the change in use of aids. In partial support of my hypothesis, scores on Exam 1 correlated with the change in the use of summary sections, \( r(208) = .16, p < .05 \). Students who did not do well on the early exam did use the summary sections more for the final than they did for the first exam. Overall, however, the use and perceived helpfulness of the pedagogical aids at Time 2 did not relate to Exam 1 scores. Unfortunately, this single small correlation does not lend itself well to any generalizations or discussion of how use changes.

Studying Change

Paired sample t tests tested whether aid usage and perceived helpfulness changed over the semester. The use of boldface, \( t(200) = 2.75, p < .01 \), and italics, \( t(197) = 2.07, p < .05 \), and the perceived helpfulness of the quizzes, \( t(197) = 5.24, p < .001 \), changed over the semester. The direction of change was counter to what I expected. The use of both types of terms dropped, and students rated the quizzes as being less helpful over time. Time 1 use and helpfulness ratings correlated moderately with Time 2 use and helpfulness ratings: average \( r(216) = .35, p < .01 \).

Few demographic differences existed. There were significant sex differences in changes in perceived helpfulness of bold, \( F(1, 214) = 6.81 \); italics, \( F(1, 214) = 4.34 \); key terms, \( F(1, 214) = 4.11 \); and practice questions, \( F(1, 214) = 5.08 \), all \( p \) values < .05, Wilk’s \( \lambda F(5, 211) = 2.70, p < .05 \). For each, women’s perceived helpfulness ratings dropped over the semester whereas men showed an increase in ratings of helpfulness.

Studying Use and Helpfulness

Consistent with Gurung (2003), the reported use and the perceived helpfulness of the different pedagogical aids were not significantly associated with specific exam scores (partial correlations controlled for ability). Furthermore, the use of a mean score for all four exams similar to Gurung (2003) showed that the perceived helpfulness of key terms was negatively correlated with exam performance. Some pedagogical aids can be dangerous detours.

The use of pedagogical aids at Time 1 and Time 2 varied with sex. For example, at Time 1 there was a significant multivariate test for use of aids, Wilk’s \( \lambda F(5, 229) = 2.64 \). Women used bold terms, \( F(1, 215) = 6.13 \), and italics, \( F(1, 215) = 4.84 \), more than men did (all \( p \) values < .05). Results for Time 2 were similar. There were no sex differences in perceived helpfulness at either time.

More able students (as assessed by ACT scores and high school GPA) used key terms, \( r(223) = -.24, p < .05 \), practice questions, \( r(223) = -.15, p < .05 \), and summaries, \( r(223) = -.19, p < .01 \), less, suggesting an inverse relationship between academic ability and use of learning enhancers. These findings varied somewhat from Weiten et al. (1996), who found that students’ educational experience, as measured by their year in school and their previous academic grade point averages, did not relate strongly to their ratings of pedagogical aids.

In future studies a separate measure of aid use and type of studying done (e.g., reading, using flash cards, practicing applications of material with a partner) will be critical to understanding how aid use relates to exam scores. I tested the use of one aid. I used an ANCOVA (controlling for ability and effort) to test whether there is a “best way” to take the online quizzes. I compared the students who read the chapter and then took the quiz with students who just took the quiz with the book open, who guessed, or who copied. Students who read the chapter before taking the quiz had significantly higher exam scores, \( F(1, 189) = 4.35, p < .01 \), scoring a mean of 81 points compared to students who took it without reading the chapter first, who scored 74.

Strengths and Limitations

There are many strengths to this study that lend credence to its implications. I controlled for two factors that account for a large amount of variance in exam performance, namely...
student ability and student effort or time studied. In addition, the prospective data collection allowed me to assess change over time. Nonetheless, there are limitations to this study. Most important, the assessments may have been too far apart to actually capture change in use of pedagogical aids. Students who performed poorly on the first test might actually have increased their use of the pedagogical aids for the second test. If they still continued to do badly, they might then have stopped using the aids by the third and fourth tests. By assessing usage only after the final, I was not able to examine such curvilinear trends in usage and perceived helpfulness. Second, the potential for impression management could contaminate responding. Third, I measured effort with only a single item, and a better assessment of how students study would be preferable (Cross & Steadman, 1996). Finally, it is also possible that these findings may vary for texts of different difficulty levels.

Conclusions

Doing poorly on an exam did not have any effect on how much students used pedagogical aids or how helpful students rated them after that point. Furthermore, the use of aids in general did not change substantially over a semester. Student study habits were relatively stable even in the face of evidence that what they were doing was not adequate. Students’ ability to predict when they need to learn more, based on their judgment of what they already know, is a life-long learning tool. Pedagogical aids, if designed well and used properly, can help students get a better sense of their knowledge levels and help guide further study. Students who do not do well on an exam should increase (or improve) their use of aids to better prepare them for future exams. Unfortunately, the results of this study suggest that a change in aid use did not take place. Pedagogical aids could be learning enhancers if students used them to find out what they know and to guide studying.

My use measures did not assess how students used the aid. Did students just memorize the key, bold, and italic terms, or did they test whether they knew what the terms were? Did they memorize the answers to the online quizzes and practice questions, or did they actually test themselves? Research designed to answer these questions is critical. When students know the level of their knowledge, only then can they effectively self-direct their learning of new material. Instead of simply advertising the number of aids in a book or directing students to use them, instructors should stress the best ways to do so.

References


Notes

1. The University of Wisconsin System Teaching Fellows Program and the Office of Professional and Instructional Development (OPID), and the University of Wisconsin, Green Bay, Department of Human Development funded this research.
2. I thank Krista Olearnick for her insights on metacognition, Randolph Smith and three anonymous reviewers for their comments, and April Hill and Danielle Edens for helping collect the data.
3. Send correspondence to Regan A. R. Gurung, Department of Human Development and Psychology, University of Wisconsin, Green Bay, 2420 Nicolet Drive, MAC C318, Green Bay, WI 54311; e-mail: gurungr@uwgb.edu.
We experimentally investigated the effect of required textbook study guides on undergraduates' multiple-choice exam performance for introductory psychology. Students in 2 sections participated. Students required to complete the study guide performed significantly better than students in the control section. Students who reported completing 75% or more of the study guide did not perform significantly better than students who completed 25% or less. Results offer support for the effectiveness of study guides in courses that use multiple-choice exams; however, more exercises may not enhance performance compared to fewer exercises. Students who used the study guide had a positive perception of the usefulness of the study guide, and most stated that they would voluntarily use one in the future.

College-level psychology courses commonly include textbook study guides, either as a requirement for the course or as an optional study aid. Despite their prevalence, there is little evidence about their effectiveness. Given the dearth of empirical evidence to guide instructors' decision-making process in whether to assign the textbook study guide, instructors are left to their perceptions or anecdotal evidence regarding the usefulness of these learning aids. This lack of evidence undermines efforts to teach scientific psychology by putting instructors in the position of endorsing textbook study guides without empirical evidence. In addition, many students spend significant amounts of money, time, and effort completing study guides without evidence that they enhance performance. Gurung (2003) argued that the misuse or overuse of study aids can hinder learning and performance, urging instructors to pay close attention to the effectiveness of pedagogical aids. He found that students perceived practice test questions and text fonts (bold or italic key terms) as useful despite the lack of a significant positive correlation between the students' use of these study aids and exam performance. This finding emphasizes the potential discrepancy between perceived and actual effectiveness of materials such as study guides, strengthening the case for empirically investigating the effects of these common study aids on student performance. In an effort to provide evidence, we experimentally investigated the effect of required textbook study guides on student learning in introductory psychology as assessed through multiple-choice exams.

There are a priori reasons to predict that textbook study guides would have a positive effect on learning. For one, these guides offer students an opportunity to effortfully process and manipulate course material (Craik & Lockhart, 1972), typically via fill-in-the-blank questions, matching questions, multiple-choice questions, and essay questions. Furthermore, completing these exercises encourages students to apply information to examples and to focus their study efforts on understanding the meaning of key concepts rather than on mere memorization. Most study guides also provide explanations for answers to promote students' understanding of the material by providing immediate feedback. Finally, the multiple-choice component of study guides may maximize transfer-appropriate processing in courses with a large multiple-choice test component. According to transfer-appropriate processing theory (Morris, Bransford, & Franks, 1977), retrieval of information is enhanced if it requires the same kind of mental processing as used during studying, implying that multiple-choice study guide questions should strongly enhance performance on multiple-choice exams (Herrmann, Raybeck, & Gutman, 1993). The consistency between the mental processes used during
Study and testing maximizes the effectiveness of retrieval cues, thereby maximizing recall (Ellis & Hunt, 1993).

Surprisingly, given the prevalence of study guides as supplementary material for introductory psychology texts, we found only three experimental studies that assessed the effect of published study guides on course grades, including traditional bound paper versions and computer versions (e.g., Web-based activities or interactive CD-ROMs). Reittinger and Crowley-Long (1992) compared final grades for an introductory psychology course between students who voluntarily completed a bound paper study guide that accompanied their textbook and students who chose not to complete the study guide regardless of whether they had been assigned the study guide. Despite the nonsignificant difference in final grades between the study guide and nonstudy guide students, Reittinger and Crowley-Long reported that an overwhelming majority of the study guide students found the study guides to be helpful and would opt to use them again given the opportunity. Flora and Logan (1996) required students in one section to complete a computerized study guide that accompanied their textbook for the second, fourth, and fifth exams, whereas students in a different section completed the computerized study guide for the first, third, and fifth exams. Student performance on the exams for which the computerized study guide was required was significantly higher than student performance on the exams without the computerized study guide requirement. Because students could complete the computerized study guide for all of the exams, Flora and Logan concluded that the observed differences between groups would have been even larger in a more precisely controlled study. Similar to Reittinger and Crowley-Long’s findings, the majority of students in Flora and Logan’s study thought the study guide was useful and recommended requiring it in future introductory psychology courses. Over three semesters, Daniel and King (2003) compared the effect of study guides, Web-based activities (multiple-choice questions, crossword puzzles, flash cards, etc.), and Web-based multiple-choice quizzes on multiple-choice exam performance. Student exam performance was significantly better for the online multiple-choice quiz group compared to the study guide and Web activities groups.

Addressing Limitations of Previous Literature

This study improved on previous research in several important ways. First, we examined two different sections of introductory psychology: Students in one section were assigned the study guide whereas students in the other section were not assigned the study guide. This between-groups design offsets potential problems such as diffusion effects that result from within-class experimental designs (Craven, Marsh, Debus, & Jayasinghe, 2001), related to differential requirements within a section and across the semester (Flora & Logan, 1996). It also maximizes external validity by ensuring consistent scholastic expectations in the same way that a typical course would (as opposed to changing expectations over the course of a semester). Second, we systematically manipulated the mandatory study guide assignment between sections to minimize the potential confound of self-selection factors, yet we allowed any student to voluntarily complete the study guide so that we could measure the performance of students who self-selected to complete the study guide (Reittinger & Crowley-Long, 1992). Related to this issue, we assessed self-reported completion rates for students in both sections to capture the degrees of completion effect for the required and nonrequired study guide on exam performance. We were able to ascertain whether completing more of the study guide exercises led to greater improvements compared to completing none or few of the study guide exercises. Additionally, concurrent sections taught by the same instructor controlled for instructor characteristics (personality, enthusiasm, likability, knowledge, etc.) and provided better equivalence across sections than designs comparing the same course over successive semesters or using a two-instructor approach (Daniel & King, 2003; Reittinger & Crowley-Long, 1992).

Outcome variables for pedagogical research can be either objective or subjective. In addition to assessing student perceptions regarding the utility of study guides, we chose multiple-choice exams as our objective measure of student performance because of their standardized nature and their prevalence in large survey introductory psychology courses. To summarize, our primary research question focused on whether requiring a textbook study guide affected student multiple-choice exam scores compared to not requiring a textbook study guide.

Method

Participants

Two hundred thirty-six undergraduates from a midsized southwestern state university participated. There were 113 students in the study guide section and 123 students in the control section. There were no significant differences between the sections with regard to gender, class status, ethnicity, or age. Students in the study were 69% women; 57% freshman; 80% White, 9% Hispanic, 6% Native American, and 3% other (Asian, African American, international, and unknown); and ranged in age from 17 to 60 (M = 20.6 years). Chi-square analyses revealed no significant differences between sections with respect to high school grade point average (GPA), college GPA, Scholastic Aptitude Test (SAT) scores, and American College Test (ACT) scores (p > .05).

Materials

Study guide. We required the entire study guide (Proc-tor, 2000) that accompanies Psychology: The Adaptive Mind (Nairne, 2000) in one section. Specific study guide activities included learning objectives, vocabulary, fill-in-the-blank concepts, multiple-choice questions, short essay questions, matching, and language enhancement. The learning objectives (M per chapter = 30, range 26 to 33) provided a condensed summary of the chapter by highlighting major points. An average of 37 vocabulary words (range = 21 to 69) were in each chapter, highlighting text definitions. Students could practice recalling text definitions and basic conceptual material from the chapter for the fill-in-the-blank exercises (M per chapter = 53, range 36 to 64). An average of 61 multiple-choice questions (range =
45 to 76) were in each chapter to test student mastery of the factual and conceptual information, text definitions, and application of text material. The short essay questions (M per chapter = 6, range = 4 to 8) required students to recall information from the chapter. The matching exercises (M per chapter = 15, range = 12 to 16) required students to link terms with their definitions. Each chapter of the study guide included two pages of language enhancement exercises designed to boost general language comprehension skills.

Subjective questionnaire. In addition to demographic information, we asked participants to rate how much of the study guide they completed (25%, 50%, 75%, 100%, or did not complete study guide), whether they thought the study guide improved their grade (yes, no, or did not do study guide), and whether they would opt to use the study guide in future courses given the opportunity (yes, no, or did not do study guide).

Four examinations. Students in the two sections completed four identical multiple-choice exams (50 questions each) derived from the textbook test bank. Of the 200 exam questions, 38% of the questions tested mastery of factual information contained in the text (e.g., experimental outcomes), 29% tested mastery of key terms, 14% focused on extending the text material to new situations, 12% focused on textbook examples to illustrate key concepts, and 7% assessed deeper understanding and integration of key concepts. An average of 22% of the exam questions were also in the study guide.

Course and instructor evaluations. Participants completed the college’s standard course and instructor evaluation questionnaire. The quantitative portion included an assessment of course organization, clarity of presentation, degree to which the course increased understanding, openness of classroom climate, clarity of course requirements, degree of challenge to think critically about subject, fairness of grading, relevance of assigned readings, availability of instructor, and effectiveness in the use of examples and illustrations. The response format ranged from 1 (unacceptable) to 5 (excellent).

Design and Procedures

Students in two introductory psychology sections taught by the third author participated in this study. In one section, the entire textbook study guide (Proctor, 2000) was required. Study guides were graded on a 4-point scale pertaining to the percentage of completion (< 25%, 25% to 50%, 50% to 75%, > 75%) after each exam (answers were provided at the back of the study guide). The study guide was worth 8% of the final course grade. The study guide was not assigned to students in the second section. These students had the option to complete the study guide: It was not recommended nor objectively assessed for completion. Student grades in this section were based solely on exam performance.

Both sections had the same textbook, course policies (with the exception of the study guide requirement), outlines, content, and exams. We collected student evaluations of the instructor and course 2 weeks prior to the end of the semester as required by university policy. Immediately following the final exam, students in both sections voluntarily completed questionnaires that included demographic information, as well as use of and perceived usefulness of the study guide.

Results

We used an alpha level of p = .05 for all statistical tests. Our analyses focused on comparisons between the experimental and control sections and between individuals who completed the study guide and individuals who did not complete the study guide, regardless of course requirement, for average exam performance and individual exam performance. We also examined student perceptions of the study guide.

Student Performance

Mean exam performance for all participants was 76.57 (SD = 12.15). Mean exam performance was 77.96 (SD = 12.04) for the study guide section and 75.16 (SD = 12.13) for the control section. We conducted an ANCOVA with section as the independent variable, exam performance as the dependent measure, and high school GPA as a covariate (we excluded 1 participant who did not provide high school GPA). The difference between the study guide and control sections was significant, F(1, 232) = 4.19, p = .04, η² = .02. High school GPA was a significant covariate, F(1, 232) = 12.03, p = .001, η² = .05.

Subsequent analyses focused on comparisons between individuals who reported completing 75% or more of the study guide (completers) and individuals who reported completing 25% or less of the study guide (noncompleters), regardless of course requirements. We excluded students who reported completing 50% of the study guide from these analyses. We used self-reported completion because the students in the control group did not hand in study guides for an objective assessment of completion. In the study guide section, we categorized 86 students as completers and 15 students as noncompleters. In the control section, we categorized 15 students as completers and 95 students as noncompleters. High school GPA was not significantly different for completers and noncompleters, χ²(3, N = 210) = .90, p = .83. There was a trend in the direction of better exam performance for completers: Mean exam performance was 78.3 (SD = 12.3) for completers and 75.6 (SD = 12.1) for noncompleters. However, an ANCOVA with high school GPA as the covariate, completion of the study guide as the independent variable, and exam performance as the dependent measure revealed that the main effect of completion was not significant, F(1, 207) = 2.96, p = .087, η² = .01 (as earlier, we excluded 1 participant for not providing high school GPA). High school GPA was a significant covariate, F(1, 207) = 8.85, p = .003, η² = .04.

We then analyzed performance on each of the individual exams. The greatest disparity in performance occurred for the first exam, with mean scores of 79.8 for the study guide section and 74.7 for the control section. The difference in performance was almost as great for the last exam, with mean scores of 81.7 for the study guide section and 78.3 for the control section. Differences on the middle two exams were
smaller (both differences in mean scores < 2 percentage points). We used ANCOVAs with high school GPA as a covariate, excluding the 1 participant who did not provide GPA, to analyze the difference in scores across sections for each exam. The difference between sections was significant for Exam 1, \( F(1, 232) = 10.88, p = .001, \eta^2 = .05 \); and for Exam 4, \( F(1, 232) = 4.70, p = .031, \eta^2 = .02 \); but not for Exam 2, \( F(1, 232) = .23, p = .63, \eta^2 < .01 \); or for Exam 3, \( F(1, 232) = 1.20, p = .28, \eta^2 < .01 \).

Student Perceptions

We assessed student perceptions of the course and the study guide because these perceptions could be an important factor in instructor decisions to assign the tool. Mean course evaluations on the 5-point scale were identical for the two sections (4.6). Requiring the study guide did not affect student evaluation of the course.

We addressed specifically how students perceived the study guide, in particular whether students thought that the study guide improved their grades and whether students would voluntarily complete a study guide in the future. Excluding those students who responded “did not do study guide,” 74% of students in the study guide section and 73.5% of the students in the control section responded that the study guide did improve their grades, and 59.2% of the students in the study guide section and 87.9% of the students in the control section stated they would voluntarily complete a study guide in the future. Thus, students who used the study guide had a generally positive perception of the usefulness of the study guide and were, to some extent, accurate in their perception that the study guide was useful. Note that a large proportion of students assigned the study guide endorsed the idea that the guide improved their grades, yet a smaller proportion stated willingness to complete one in the future. The students who willingly completed the study guide when it was not required were much more likely to willingly use one again in the future.

Discussion

Student Performance

Students in the required study guide section performed significantly better than students in the control section. The effect size was small, yet statistically significant. This difference is unlikely to have resulted from overall differences in academic preparation between the sections given that we statistically controlled for high school GPA as an indication of academic preparation. Four possible mechanisms may have contributed to this pattern of results. First, the study guide that accompanies the textbook may simply have facilitated learning the textbook material just as it was designed to do, namely through effortful processing, practice, and application of concepts in novel ways. Second, the finding that the effect of the study guide was greatest for the first and last exams suggests that the study guide may have functioned as a mechanism that encouraged the student to become engaged in the course material earlier in the semester (study guide completion was assessed each exam day) and remain engaged with the course material through the final exam of the semester. That is, the study guide exposed students to the course material and similar types of questions that will appear on the exam prior to the examination. As we required students to complete the study guide and turn it in on the day of the final exam, they were “forced” to work with the material until the end of the semester. The many pressures that emerge toward the end of the semester may have more easily distracted students in the control section, who did not have the ongoing requirement of completing study guide exercises. A third possible mechanism is that practicing multiple-choice questions in the study guide enhanced students’ test-taking skills and reduced test anxiety, rather than, or in addition to, increasing their learning of the material. A fourth possible mechanism is that students were exposed to 22% of the exam questions prior to the exams due to the overlap of study guide and exam questions. Given that there were an average of 61 multiple-choice questions per chapter, and thus an average of 244 multiple-choice questions in the study guide for each exam, it seems unlikely that students improved their exam scores by memorizing the specific answers for specific questions they had previously seen. Although our view is that this is not the mechanism by which students improved performance, we do think that the potential overlap of questions in the study guide and the exam may actually motivate students to complete the study guide.

One may question the practical significance for students of a 2% increase in exam performance. In other words, are students likely to expend a significant amount of time and effort for a mere 2%? On one hand, we have seen students eagerly participate in extra credit activities for fewer points. On the other hand, it could be that students in the control section were spending the same amount of time and effort in less productive types of studying. Future research should assess the study time and effort trade-off with grade improvement.

Subsequent analyses raised some interesting issues regarding the relationship between study guide completion and student performance. Across both sections, students who reported completing 75% or more of the study guide had slightly higher exam scores than those who reported completing 25% or less. Although the trend was in the direction favoring the study guide completers, the trend was not significant. One possible reason for the lack of significant results is that more may not necessarily be better when it comes to the study guide exercises. Some of the exercises may be superfluous, as they do not relate directly to the course material (e.g., language enhancement) or to the type of questions that appear in the exam (e.g., fill-in-the-blank, short essay, matching). It may be that the multiple-choice questions account for the improved exam performance as suggested by the transfer-appropriate processing theory (Herrmann et al., 1993): Retrieval of information is enhanced if it requires the same kind of mental processing as used during studying. This interpretation is supported by Daniel and King’s (2003) finding that student exam performance was significantly better only for students who completed online multiple-choice quizzes, not for students who completed the study guide or a wide range of Web activities such as essay and multiple-choice questions, crossword puzzles, and flash cards. A common assumption, held by students and instructors alike,
is that if a student spends time “studying” course material, the student will learn the information and perform well on the exams. In all likelihood, not all study time is equal; time spent learning must be actively used and students need to engage in productive activities that correspond to the testing method. In a similar study currently underway (Dickson, Miller, & Devoley, 2004), we found that students who selectively use the study guide are most likely to complete the multiple-choice, matching, fill-in-the-blank, and vocabulary sections, and least likely to complete the language enhancement and essay sections. Students may have completed the study activities they perceived as useful, such as multiple-choice questions and vocabulary, and then engaged in self-initiated study strategies rather than “wasting” their study time on the required study guide activities that either required more work, such as the essay questions, or that did not correspond highly with the testing method (e.g., essay questions and language enhancement). This selective completion strategy may have diluted the effect of the study guide on test performance. Future research should examine the effectiveness of specific study guide exercises on multiple-choice exam performance.

The discrepancy between the significant section effect and the nonsignificant effect for completers compared to noncompleters highlights the need to further explore potential differences between students who voluntarily complete the study guide and those who do not. Future research needs to examine motivational and psychosocial factors that differentiate students who elect to complete the study guide and the varying degrees of effort put forth toward completing these exercises. The finding that only 14% of the students in the control section completed 75% or more of the nonrequired study guide suggests that the vast majority of students will not opt to complete “extra” work. On the other hand, the finding that 15% of students in the experimental section completed 25% or less of the study guide suggests that individual motivation is still an important factor determining student use of study guides, regardless of course requirements. To summarize, requiring the study guide is one way for instructors to improve their students’ performance on multiple-choice exams. At the same time, instructors need to realize that a fairly substantial minority of students will refuse to complete such aids even if there is a grade penalty.

**Student Perceptions**

Our primary focus was on the effect of the study guide on student performance, but we also examined student perceptions because student attitudes toward and perceptions of the required work could potentially affect the classroom environment as well as overall perceptions of the course and instructor. The good news is that we obtained identical course and instructor evaluations in both sections, demonstrating that the difference in requirements and workload did not alter students’ perceptions of the course or the instructor. On the other hand, we found that students assigned to complete the study guide were somewhat less willing to voluntarily complete a study guide in the future, despite the fact that many of students in this section stated that they thought the study guide helped improve their grades. Many students seem to hold the contradictory beliefs that the study guide improved their grades but that it would not be worth doing in the future. Perhaps the promise of an improved grade is not compelling enough for students to voluntarily commit themselves to study such effortful and time-consuming exercises. This possibility further encourages instructors to make these study guides a required, graded component of courses rather than a completely optional assignment, as the latter approach is unlikely to result in substantial student participation or gains in learning.

**Conclusions**

This study supports the effectiveness of a published study guide in an introductory psychology course that uses multiple-choice exams. We acknowledge, however, that the fact that it examined the effectiveness of only one published study guide, used in conjunction with only one textbook, limits its generalizability somewhat. Our impression is that the Nairne (2000) textbook and study guide are reasonably representative of the spectrum of materials available for introductory psychology. However, further research is needed to establish that these findings do extend to other published introductory psychology materials. In addition, future research should examine the effectiveness of specific study guide exercises on student performance. The completers versus noncompleter analyses indicated that more is not necessarily better; thus, additional research should also assess whether requiring students to complete only multiple-choice study guide questions differentially affects performance on multiple-choice exams compared to assigning a variety of study guide exercises as required in this study. Because computerized study guides are increasingly available to students free of charge via textbook Web sites, future research should also examine the effectiveness of study guide material delivered via computer.

**References**


Notes

1. Portions of this article were presented at the meeting of the Rocky Mountain Psychological Association, Denver, CO, April 2003.
2. We thank the anonymous reviewers for their insightful comments.
3. Send correspondence to K. Laurie Dickson, Department of Psychology, Northern Arizona University, P.O. Box 15106, Flagstaff, AZ 86011; e-mail: laurie.dickson@nau.edu.
Students took four multiple-choice question exams (50 questions each) derived from the textbook test bank and based on the learning objectives. Seventy-six test items required lower order skills, and 124 test items involved higher order skills. (For definitions of these terms, see Dickson & Miller, 2005.)

Results

To examine the effect of self-constructed versus other-constructed crib cards on exam performance, we computed each student’s average score for self-constructed crib card exams ($M = 78.54$, $SD = 9.03$) and other-constructed crib card exams ($M = 76.00$, $SD = 8.49$). Students performed better with the other-constructed cards than with the self-constructed cards, $t(31) = 4.16, p < .001$.

Discussion

Exam performance was significantly better when students unexpectedly used a crib card constructed by someone else. A likely explanation for this finding is that when students do not expect to use a crib card, they make a stronger effort to understand the material prior to the exam. In contrast, when they expect to have information available on a crib card during the exam, they do not study the material as effectively. Thus, the unexpected crib card contributes to their higher scores by serving as a memory aid or as a resource for information that they did not learn. In other words, students simply use the information on their own crib cards to answer the exam questions without learning the material. It is possible that students actively process test-related information during crib card construction, but even if so, this active processing does not compensate for the overall reduction in learning associated with self-constructed crib cards.

Another possible explanation for our results is that the quality of the other-constructed crib cards was significantly better. This possibility seems unlikely given that a student who had not taken the course constructed the crib cards with the textbook and the learning objectives, just as the students in the class did. Even if the other-constructed cards contained more information, the finding that students performed better with the premade crib cards than with their own contradicts the coding hypothesis (Dorsel & Cundiff, 1979; Hindman, 1980; Trigwell, 1987) because it suggests that having information on the crib card helps more than the process of constructing the crib card.

A related issue is that average students may not learn how to effectively engage in the coding process while constructing a crib card. In other words, they may not know what information to put on the card or how to study the information on the crib card, deficits that highlight weaknesses in student study habits in general. The fact that our research assistant, an above-average student, had not taken the course or seen the exams, yet produced a crib card that led to higher exam scores, supports this idea. Future research should examine the coding hypothesis with an intervention study by teaching students how to effectively create and use a crib card.

The construction of crib cards did not facilitate student learning, and the use of self-constructed crib cards did not enhance exam performance in this content-driven upper division psychology course. Given the mounting evidence against crib card effectiveness (Dickson & Miller, 2005; Hindman, 1980; Whitley, 1996), we recommend against the use of crib cards during exams, until teaching scholars more fully understand how crib cards affect learning and exam performance after instructors have coached students on the processes of actively selecting, organizing, and representing the information in a condensed form.

References


Notes

1. We thank the anonymous reviewers for their insightful comments and Carl Moncher for making the other-constructed crib cards.
2. Send correspondence to K. Laurie Dickson, Department of Psychology, Northern Arizona University, P.O. Box 15106, Flagstaff, AZ 86011; e-mail: laurie.dickson@nau.edu.

Effect of Study Guide Exercises on Multiple-Choice Exam Performance in Introductory Psychology

K. Laurie Dickson, Michael S. Devoley, and Michelle D. Miller
Northern Arizona University

We experimentally investigated whether requiring completion of only multiple-choice study guide questions differentially affected multiple-choice exam performance compared to requiring a variety of study guide exercises (learning goals, vocabulary, fill-in-the-blank concepts, multiple-choice questions, short essay questions, matching, and language enhancement) for introductory psychology students ($N = 244$). There was not a significant difference in multiple-choice exam performance. We conclude that requiring a variety of study guide exercises produces no significant benefit for multiple-choice exam performance over and above those produced by requiring only multiple-choice exercises.

Published study guides that accompany psychology textbooks potentially enhance student learning and exam perfor-
performance. Completing study guide exercises requires a significant investment of student time and effort, raising the issue of how to maximize returns on this investment. In particular, it is unclear whether all of the exercises typically included in a study guide produce measurable gains in exam performance. It may be the case that only certain exercises, for example the multiple-choice practice questions, actually raise multiple-choice exam performance.

Students perceive study guides that accompany their textbooks as helpful (Dickson, Miller, & Devoley, 2005; Reittinger & Crowley-Long, 1992). However, research regarding the effect of study guides on exam performance is mixed. Reittinger and Crowley-Long (1992) found nonsignificant differences in final grades for an introductory psychology course between students who voluntarily completed a textbook study guide and students who chose not to complete the study guide. Daniel and King (2003) compared the effect of required study guides, Web-based activities (multiple-choice questions, crossword puzzles, flash cards, etc.), and Web-based multiple-choice quizzes on multiple-choice exam performance. Student exam performance was significantly better for the online multiple-choice quiz group compared to the study guide and Web activities groups. Dickson et al. (2005) found that students with a study guide requirement performed significantly better than students without a study guide requirement. Results offered support for the effectiveness of required study guides in courses that use multiple-choice exams; however, the design of the Dickson et al. study did not allow the researchers to address the benefits of specific types of study guide exercises.

Given that Dickson et al. (2005) previously found that requiring a variety of study guide exercises led to higher exam scores, we assessed whether requiring only multiple-choice study guide questions differentially affected multiple-choice exam performance compared to requiring a variety of study guide exercises (learning goals, vocabulary, fill-in-the-blank concepts, multiple-choice questions, short essay questions, matching, and language enhancement). In addition, we examined student perceptions of the study guide exercises.

Method

Participants and Procedures

Two hundred forty-four students in two introductory psychology sections participated (122 students in each section). There were no significant differences between the sections with regard to gender (65.1% women) or class status (73.1% freshmen). Chi-square analyses revealed no significant differences between sections with respect to high school grade point average (GPA), college GPA, Scholastic Aptitude Test (SAT) scores, or American College Test (ACT) scores (\(p > .05\)).

The second author taught both sections with the same textbook (Nairne, 2003), course policies (with the exception of the study guide requirement), outlines, content, and exams. In one section, the instructor required all of the study guide exercises (Proctor, 2003), including learning goals, vocabulary, fill-in-the-blank concepts, multiple-choice questions, short essay questions, matching, and language enhancement (see Dickson et al., 2005). For the other section, he required only the multiple-choice exercises in the study guide. He graded the study guides for completion after each exam (8% of the final course grade). Students completed four identical multiple-choice question exams (50 questions each) derived from the textbook test bank (see Dickson et al., 2005). Immediately following the last exam, students reported their demographic information (gender, class status, high school GPA, college GPA, SAT scores, and ACT scores) and their perceptions (negative, somewhat negative, neutral, somewhat positive, positive, not applicable) regarding the usefulness of the study guide and the specific exercises.

Results

Student Performance

Mean exam performance was 75.05 (SD = 12.38) for the multiple-choice-only section and 74.23 (SD = 10.95) for the entire study guide section. We conducted an ANCOVA with section as the independent variable, exam performance as the dependent measure, and high school GPA as a covariate (we excluded 8 participants who did not provide high school GPA). High school GPA was a significant covariate, \(F(1, 233) = 18.44, p < .001, r = .27\). The difference between the multiple-choice-only and the entire study guide sections was not significant, \(F(1, 233) = .28, p = .60, r = .03\).

Most participants (69.6%) reported either positive or somewhat positive feelings about the study guide, 20.4% were neutral, and 7.8% were either negative or somewhat negative toward the study guide (2.2% missing values). When questioned about the usefulness of the various exercises in the study guide, more students (79.6%) thought positively of the multiple-choice questions than the other exercises (45.5% for true–false, 40.5% for vocabulary, 33.3% for learning goals, 21.1% for essay questions, 21.9% for language enhancement). Most participants (76.4%) thought that the multiple-choice exercises improved their exam scores, 16.9% did not think that they helped, and 5.9% reported that they did not complete the exercises (0.8% missing values).

Discussion

We did not find a significant difference in multiple-choice exam scores for students with the entire study guide requirement and students with the multiple-choice requirement. Previous research found that students who were required to complete the entire study guide performed significantly better than students in a control section and that when the study guide exercises were not required most students did not complete them (Dickson et al., 2005). Although study guide exercises led to higher multiple-choice exam scores, our findings suggest that requiring a variety of study guide exercises does not enhance multiple-choice exam performance more than requiring fewer, more relevant exercises (i.e., multiple-choice questions).

Most participants reported positive feelings about the study guide and believed that the multiple-choice exercises
improved their exam scores. Although the true–false, vocabulary, learning goals, essay questions, and language enhancement exercises were not “dangerous detours” (Gurung, 2004) because they did not hinder student performance, they also did not enhance multiple-choice exam performance beyond the benefit of the multiple-choice exercises. The congruence between the type of study guide exercise and exam question may explain the benefit. The transfer-appropriate processing theory (Herrmann, Raybeck, & Gutman, 1993) states that retrieval of information during testing is enhanced if it requires the same kind of mental processing as used during studying (Ellis & Hunt, 1993). However, the possibility exists that the principle of diminishing returns applies. Perhaps any 60 exercises per chapter are sufficient and additional investment results in relatively little payoff. For example, do 60 true–false exercises per chapter show a similar effect on multiple-choice exam performance? A comparison of varying numbers of multiple-choice exercises is warranted to address whether there is a threshold of multiple-choice exercises for enhanced multiple-choice exam performance. In addition, the effectiveness of the multiple-choice exercises could differ depending on how students use them and whether they are mandatory (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Brothen & Wambach, 2001; Grimstad & Grabe, 2004; Thomas & Rohwer, 1986). Requiring study guide exercises of any type may force students to study, which leads to improved exam performance. Is there a difference in the effectiveness of the multiple-choice exercises if students use the exercises to become familiar with the test material or to test their understanding of the material after they have studied?

There was not a significant difference in multiple-choice exam performance between students with the multiple-choice exercises requirement and the entire study guide requirement. Thus, we conclude that students in this introductory psychology course maximized their investment of time and effort by focusing on multiple-choice exercises when preparing for a multiple-choice exam, rather than additionally investing in a variety of study guide exercises.

References


Notes

1. We thank the anonymous reviewers for their insightful comments.

2. Send correspondence to K. Laurie Dickson, Department of Psychology, Northern Arizona University, P.O. Box 15106, Flagstaff, AZ 86011; e-mail: laurie.dickson@nau.edu.

An Assignment to Help Students Learn to Navigate Primary Sources of Information

Andrew N. Christopher and Mark I. Walter
Albion College

Being able to apply methodological and statistical concepts is a goal that many teachers of psychology may have for their courses. Toward this end, we describe an assignment that involves using primary source journal articles and that leads students to further their understanding of material presented in methodology and statistics courses. Student performance on a small number of exam questions suggested pedagogical value in using this assignment. We present cautions in implementing the assignment and provide resources that teachers can use.

Among the many goals that teachers of statistics and research methods establish is an increased ability by their students to understand information in primary sources. Indeed, 72% of undergraduate research methods courses and 22.7% of undergraduate statistics classes use primary sources (Oldenburg, 2005). Increased understanding of statistics and research methods is more likely when students complete specific assignments using such sources. We developed the assignment described herein to help students understand methodological and statistical information by using psychology journal articles.

Numerous articles have appeared in Teaching of Psychology to help teachers of statistics and research methods courses develop specific course activities and assignments. For instance, Dolinsky (2001) provided several practical and easy-to-implement active learning strategies for teaching statistics. Specifically, she discussed various ways to use SPSS to teach different topics (e.g., statistical interactions), how to implement active learning strategies (e.g., student presentations of conceptual information), and the use of writing as-
University students attending large introductory psychology classes reported their motivations and goals regarding preferences for course requirements and evaluation methods. Overall, students had expectations that were sometimes contradictory across preferences, were different across subgroups, and were unlikely to be compatible with faculty goals. Women preferred more opportunities for evaluation and desired to have less weight placed on each evaluation event. Women also preferred a grade distribution with a preponderance of higher grades, as did men who endorsed a performance goal. Only men who endorsed mastery goals preferred a truly normal grade distribution. On ratings of the weightings preferred for effort and for mastery, younger students wanted more weight placed on effort than did older students, and students with a mastery goal wanted less weight placed on effort than did those with a performance goal.

A large group of diverse majors enrolls in the introductory psychology course at most universities. Although the course serves many purposes for students, for psychology departments it often is a place to recruit majors. Although the course content, when effectively presented, should determine whether a student chooses the major, the student’s overall experience, including satisfaction with the evaluation procedures in the course, will have an impact. Unfortunately, you have probably been there: You reach the end of a course after having carefully explained the course requirements, grading system, and your expectations, and students are nonetheless upset because they are surprised and usually disappointed by their final grade. Some of the emotion may be due to a disagreement about what is important or appropriate in evaluating student performance. Although it is rightfully the faculty member’s responsibility to establish the evaluation criteria, an appreciation of student expectations and preferences might allow for the diffusion of students’ emotional responses and a happier experience for all.

Of course, in addition to differing from faculty in their preferences concerning evaluation, not all students are likely to have the same preferences for evaluation strategies. These possible differences, from faculty and among students, may explain much of the satisfaction (or dissatisfaction) students express when they receive grades. In this study, we investigated how students in introductory psychology classes preferred to be graded and assessed student subgroup differences in these preferences.

Several trends in higher education may contribute to a gap between professors’ and students’ expectations. Years of grade inflation in secondary and postsecondary schools may have established an expectation that class attendance and a good-faith effort should be adequate to generate a good grade (Landrum, 1999). In addition, if institutions and students regard universities as a provider of a service and students as consumers, then the consumers reasonably want to be satisfied with their purchase (Hartoonian, 1997). Dissatisfaction may lead to student attrition, ineffective recruiting of new majors, and lowered course evaluations.

Picard (1999) found that students taking introductory-level geology courses held high expectations for their outcomes in the course. Ninety-two percent reported that an A or B was a reasonable average grade for the course, whereas only 7% believed that a C would represent average performance. As to their expectations of their teachers, students most often picked “accommodates individual learning styles and abilities” as the most important quality and “friendly and warm” as the second-choice quality. If these findings are representative, then the seeds of discontent are sown and ready to reap.

Age, Goals, Gender, and Academic Motivation

Understanding subpopulations of students in terms of what motivates them and what they like and do not like about their educational experience pays off in terms of reduced attrition rates (Donohue & Wong, 1997) and informed classroom practices. Differences in perceived sources of stress as well as differences in outside responsibilities are apparent for different ages (Dill & Henley, 1998) and play a role in the enjoyment of the college experience.

Being an older student seems to be associated with a number of positive academic characteristics, such as positive self-image, an internal locus of control, greater receptivity to an atmosphere of formal learning (Nunn, 1994), and generally a more positive and optimistic profile (Dill & Henley, 1998; Harju & Eppler, 1997). It is not clear, however, that these more positive characteristics translate into higher grades (Bergin, 1995; Harju & Eppler, 1997).

In addition, academic expectations may vary as a function of gender. Nunn (1994) noted that women in college classes had a more external locus of control, expressed more academic anxiety, and held higher achievement orientations than did men.

These results suggest that subpopulations of students may have different perspectives on the academic evaluation pro-
cess. Are these differences related to systematic differences in preferences for evaluation strategies? It seems likely that as one’s style or goals in an academic environment change, so might views regarding a desirable evaluation plan. Although a considerable literature exists examining the relation between eventual grade and satisfaction with a course or instructor (e.g., Hocking, 1976; Marlin & Gaynor, 1989; for a different conclusion, see Johnson & Christian, 1990), the effects of course structure and evaluation strategies on student satisfaction have received less attention. This research provides data to better understand student motivations and goals and to identify possible group differences in preferences for course requirements and evaluation methods among introductory psychology students. It addresses what students consider appropriate evaluation strategies and how subgroups might differ.

Method

We recruited participants (N = 220; 57% first-year students, 26% sophomores, 17% advanced students) from large lecture sections of introductory psychology classes because we were interested in the expectations students bring with them early in their college experience. An experimenter received permission from the instructor to use the last 15 min of class to request participation and to administer the survey. Participation was voluntary and research participation credit was available to those who were eligible. Most students in the classes stayed, signed separate consent forms, and completed the survey.

Materials included a 48-item questionnaire that asked demographic questions and forced-choice questions about preferred classroom grading practices. The survey assessed issues that anecdotally seemed of interest to us, such as reasons for taking the particular course, the nature and number of preferred tasks, what components should be considered when determining a final grade and the differential weighting given those components, and how grades should be distributed. The survey also included a question asking if students ever were surprised by their final grade and why.

Results

Who Are They, and What Do They Want?

Most of the participants were of traditional undergraduate age, with a wide variety of majors represented from the arts and sciences college as well as professional colleges. The majority had come from homes in which parents had either attended or completed college. Thirty percent were men, 66% were White, and 24% were Black. Although the sample cannot be considered representative of all college populations, there were no indications that this group was atypical of introductory psychology course students in general.

Table 1 indicates student responses to several survey items. In general, the group was fulfilling a requirement by taking the course and wanted a good grade through fun or easy tasks. They preferred multiple-choice tests and hoped to be graded on a curve. In addition, on a scale of 0 (very few) to 8 (very many), students indicated a preference (M = 5.15, SD = 2.46) toward more rather than fewer evaluation opportunities.

Participants indicated what they considered to be a fair grade distribution. They thought that instructors should give roughly equal numbers of As, Bs, and Cs, with fewer Ds, and even fewer Fs. A grade of B would be modal, and A, B, and C would each make up about 30% of the grade distribution.

Participants weighted five components that could be considered when calculating a grade (mastery of material, effort, attendance, participation, and extra credit) on a scale ranging from 0 (0 to 10 points) to 9 (91 to 100 points). Participants awarded the most points to mastery (M = 4.29, SD = 2.37), then, in descending order, effort (M = 3.80, SD = 2.77), attendance (M = 2.18, SD = 2.61), participation (M = 2.00, SD = 2.30), and extra credit (M = 1.40, SD = 2.19).

When asked to indicate whether they generally get the final grade they expected versus receiving an unexpected grade, 71% said they generally get what they expected. When they are surprised by a grade, 58% are surprised because the grade is lower. The fact that 42% indicated that they are surprised because the grade is higher than they expected suggests that the discontent is not just due to sour grapes, but rather that there is confusion on the part of students as to how to anticipate their grade. Finally, the participants indicated why they thought the grade was lower than expected. Seventy-four percent believed that the professor did not sufficiently take the effort extended by the student into consideration when calculating the grade. Twenty percent believed the low grade did not accurately reflect their test scores, and 7% believed that their grade did not reflect participation or...
attendance. The next logical question seemed to be, if students want effort to be considered when determining their grade, then how well do they think professors can assess effort? The answer appears to be not at all that well. When given a scale from 0 (cannot assess accurately) to 8 (can assess with great accuracy), the mean response was 3.68 (SD = 1.99), slightly below the midpoint. Furthermore, a correlation between how heavily students believe professors should weight effort and how well they think professors can assess effort showed no relation between the two variables, r(212) = .01.

Gender × Goal × Age Group Comparisons

To examine the data more closely, we considered three ex post facto factors: gender, age, and academic goals. A rough median split produced groups of younger (under 20, n = 92) and older (20 or older, n = 128) participants. Student response to a question about what they really wanted from a class determined goal designation: A preference for receiving “a good grade” indicated a performance goal (n = 138), whereas a choice of “learning new information” indicated a mastery goal (n = 73). We validated goal identification for the entire group by doing a chi-square analysis using goal designation as one variable and preferred task as the second variable. The significant result, χ²(4, N = 216) = 47.64, p < .001, indicated that 78% of performance-oriented people selected tasks that were either fun or for which they were certain of getting a good grade, whereas 59% of mastery-oriented people indicated they would choose tasks on which they would learn something new or that would reinforce learning. Although not a perfect division, the dominant motivation certainly varied between these groups.

We next calculated a series of Gender × Goal × Age Group ANOVAs using the survey answers as dependent variables. Analysis of the answers regarding the ideal grade distribution, using grade distribution as a repeated measure, produced significant effects of distribution, F(4, 804) = 65.02, p < .001; Distribution × Gender, F(4, 804) = 7.83, p < .001; Distribution × Goal, F(4, 804) = 3.90, p < .004; and Distribution × Gender × Goal, F(4, 804) = 4.01, p < .003. The means for the final interaction appear in Table 2. Percentage categories were established as follows: An answer of 0 indicated that the participant believed that 0% of the students in a class should receive that specific grade, 1 = 10%, 2 = 20%, and so on, up to 9 = 90%.

One way to ensure dissatisfaction is to have expectations on the last variable produced a significant main effect of weighting, F(1, 202) = 9.27, p = .003. Overall, participants believed professors should put the most weight on mastery (M = 4.29, SD = 2.37) rather than effort (M = 3.80, SD = 2.77). This main effect, however, was modified by a near-significant interaction with goal, F(1, 202) = 3.61, p = .059, and a significant interaction with age group, F(1, 202) = 4.10, p = .044. Post hoc analyses of the Weighting × Goal interaction indicated that individuals with performance goals rated mastery (M = 4.36, SD = 0.24) and effort (M = 4.06, SD = 0.28) equally, but that those holding mastery goals believed the professor should weight mastery (M = 4.48, SD = 0.31) more heavily than effort (M = 3.17, SD = 0.36), t(73) = 3.30, p = .002. Further analysis of the Weighting × Age Group interaction produced significant differences between the two age groups’ weightings of effort, t(213) = 2.19, p = .03, with younger students weighting effort (M = 4.04, SD = 0.36) more heavily than older students (M = 3.19, SD = 0.27). The two age groups did not differ in the strength of weighting mastery (M = 4.31, SD = 0.31 and M = 4.53, SD = 0.23 for younger and older participants, respectively). For the younger individuals, the weighting of mastery was equal to that of effort, whereas older individuals weighted mastery more heavily than effort, t(122) = 3.56, p = .001.

A last ANOVA examined differences in the preferred number of assignments on which to base a final grade. A significant main effect of gender, F(1, 203) = 8.92, p = .003, indicated that women preferred more assignments than did men (Ms = 5.65 and 4.16 and SDs = 2.24 and 2.51, respectively).

Discussion

General Preferences

One way to ensure dissatisfaction is to have expectations that one cannot easily meet. To some extent, this situation was present in the sample examined. A characterization of

### Table 2. Mean Grade Distribution for Each Gender × Goal Subgroup

<table>
<thead>
<tr>
<th>Goal</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3.53</td>
<td>.44</td>
</tr>
<tr>
<td>B</td>
<td>3.70</td>
<td>.36</td>
</tr>
<tr>
<td>C</td>
<td>3.35</td>
<td>.34</td>
</tr>
<tr>
<td>D</td>
<td>2.18</td>
<td>.29</td>
</tr>
<tr>
<td>F</td>
<td>1.23</td>
<td>.27</td>
</tr>
<tr>
<td>Mastery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1.81</td>
<td>.52</td>
</tr>
<tr>
<td>B</td>
<td>2.81</td>
<td>.42</td>
</tr>
<tr>
<td>C</td>
<td>3.38</td>
<td>.40</td>
</tr>
<tr>
<td>D</td>
<td>2.47</td>
<td>.34</td>
</tr>
<tr>
<td>F</td>
<td>1.88</td>
<td>.32</td>
</tr>
</tbody>
</table>

Note. Percentage categories were established as follows: An answer of 0 indicated that the participant believed that 0% of the students in a class should receive that specific grade, 1 = 10%, 2 = 20%, and so on, up to 9 = 90%.
the majority of the students in this sample reveals many potential sources of conflict. Many more students in the sample wanted success (65%) as the outcome than wanted learning (35%). As a starting point, this goal is undoubtedly not consistent with most faculty expectations. Faculty usually focus on learning. The students wanted to be graded on a normal curve (83%), but they believed a “fair” distribution of grades is not at all close to “normal.” Thus, if one condition is met, the other must be violated. Students wanted their assignments to be interesting and fun or easy, yet they believed mastery of the material should be most heavily weighted. Only if developing mastery can be consistently fun or easy will they be pleased with the process. Given their strong preference for multiple-choice tests, they either must believe taking such tests is fun or they want them all to be easy. Finally, they believed instructors should reward effort, and when they get a lower grade than expected they believed failure to consider effort was a likely cause, despite their recognition that instructors cannot accurately assess effort. So, students wanted to weigh heavily something that cannot be reliably measured. Problems seem likely to result within this student-desired system.

An important qualification of this interpretation, however, is the fact that most of the students surveyed were taking a class selected either to meet a general education requirement or as a first step toward a major in psychology. Responses may have been different for smaller, more advanced classes within each person’s major. Thus, the results presented here might not generalize beyond introductory psychology classes or similar large introductory courses. Of course, for many departments, large introductory courses are a major part of the teaching workload.

Subgroup Differences

The subgroup comparisons (goal, age, and gender) produced some interesting differences, but these findings require cautious interpretation because the sample was predominantly young and female. Women preferred more opportunities for evaluation than did men, perhaps due to their greater academic anxiety (Nunn, 1994) and the desire to have less weight placed on each evaluation event. An academic anxiety argument also is consistent with women’s apparent preference for a nonnormal grade distribution with a preponderance of higher grades, although in this view they were similar to men who had a performance goal. Only the mastery-oriented men saw a truly normal grade distribution as fair. The only age difference, indicating a lowered rating of effort as important in evaluation by older students, probably reflects the realization over time that effort, as an unreliable factor, is unlikely to be a good basis for evaluation. Obviously, a replication across a wider range of ages would be desirable before placing too much emphasis on this difference, but it is not unreasonable to expect that as one gains experience in life one learns that trying but not succeeding seldom leads to significant rewards.

The final subgroup difference, between students having a mastery versus a performance goal, is interesting only because it reinforces the fact that all students recognized mastery as the most relevant criterion for evaluation. Those with a performance goal rated effort as equally important as mastery, but even they accepted that no other criterion was more important.

Overall, these students’ preferences describe a set of conditions almost guaranteed to create high levels of dissatisfaction with the outcomes of evaluation. Instructors might be well served by taking time at the beginning of a course to clarify some of the logic used in evaluation. If grades are curved, there cannot be a disproportionate number of high grades. If an instructor cannot reliably assess effort, it cannot count heavily in evaluation. If mastery is the best basis for evaluation, learning, not outcome, will have to be the focus of most assignments. Although these relations may seem obvious to instructors, they may not be as clear to students. Ending on a positive note, most students did get the grade they expected in courses, so satisfied or not, most knew what was coming.

References


Note

Send correspondence and requests for the survey instrument to either author at the Department of Psychology, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223; e-mail: jgaultny@email.uncc.edu or acann@email.uncc.edu.
Prior Knowledge and Its Relevance to Student Achievement in Introduction to Psychology

Ross A. Thompson
Byron L. Zamboanga
University of Nebraska

Educational psychologists find that prior knowledge influences new learning. We examined whether course achievement for introductory psychology students is facilitated or impaired by their prior knowledge of psychology. We administered a pretest exam to 422 students early in the semester and gathered subsequent exam scores and other measures of student achievement. Students generally performed poorly on the pretest, as expected, but regressions revealed that pretest scores were uniquely positive, significant predictors of student achievement with other influences on achievement (e.g., homework, attendance) controlled. Further analyses suggested that prior knowledge is a significant resource that faculty can enlist in their instruction.

The influence of prior knowledge on behavior is familiar to psychologists. In studies of the confirmation bias, concept development, mental sets, preattentive processing, selective perception, prejudice, and learning and skill acquisition, researchers have shown how preexisting knowledge and expectations can significantly bias how people respond to new events. The influence of prior knowledge on learning is also a concern for educators. Students are likely to be influenced by preexisting assumptions and beliefs when introduced to new information, especially in courses like psychology in which the topics are so readily related to everyday life. In this study, we examined students’ prior knowledge of psychological topics in an Introduction to Psychology course and the relevance of prior knowledge to their subsequent course performance.

Research in cognitive psychology has shown that individuals with greater preexisting knowledge about a topic generally understand and remember more than those with more limited prior knowledge (Chi & Ceci, 1987; Glaser, 1984; Schneider & Pressley, 1997). Indeed, constructivist theory argues that all new learning builds on prior understanding (see Committee on Developments in the Science of Learning, National Research Council, 1999; McCormick & Pressley, 1997). Research reviews in educational psychology have concluded that prior knowledge within a specific domain benefits students’ learning and achievement (Alexander & Judy, 1988; Dochy, Segers, & Buehl, 1999). This conclusion has been confirmed in studies of a variety of academic content domains, including physics and mathematics (Hudson & Rottmann, 1981), writing ability and text processing (McCutcheon, 1986), economics (Dochy, 1992), and computer programming (Klahr & Carver, 1988), with students ranging from elementary grades to graduate school.

At times, however, prior knowledge can make it difficult to understand or learn new information (Alexander & Judy, 1988; Committee on Developments in the Science of Learning, National Research Council, 1999; Dochy et al., 1999). Difficulty is especially likely if preexisting information is inaccurate or incomplete, such as when students generalize inappropriately from everyday experiences or from what they learn in the popular media (Chinn & Brewer, 1993; Perkins & Simmons, 1988). Although interference from prior misconceptions is most often observed with young children, mistaken assumptions and prior beliefs can also undermine college students’ learning of physics (Clement, 1982), biology (Fisher, Wandorsee, & Moody, 2000; Wandorsee, 1986), and other topics (Guzzetti, Snyder, Glass, & Gamas, 1993). Remarkably, prior beliefs may be highly resistant to change, even in the context of formal coursework (Fisher et al., 2000).

Undergraduate students in an Introduction to Psychology course arrive on the first day of class with considerable prior knowledge of psychological concepts. Prior knowledge of psychology derives from many sources, including formal coursework in secondary school and informal lessons from folk theories, the media, and everyday experience. These prior beliefs may facilitate student learning of psychological concepts, but prior knowledge can also impair students’ understanding. For example, discussions in the popular media of brain development, hemispheric specialization, psychological disorders (e.g., depression, “multiple personality”), and hereditary influences on development are often good reflections of the state of scientific knowledge, but they are also usually incomplete and can exaggerate, obscure, or misrepresent current knowledge (Thompson & Nelson, 2001). Folk theories of interpersonal attraction inaccurately teach that “opposites attract” but also instruct, consistent with social psychological research, that “birds of a feather flock together.” Consequently, prior understanding may facilitate students’ comprehension of these topics but may simultaneously impede clear and accurate understanding.

Although research psychologists carefully assess prior knowledge in experimental investigations, using a pretest–posttest design to assess (and sometimes control for) previous understanding, it is rare for psychology class instructors to do so. Yet there are several reasons why assessments of prior knowledge, conducted through a pretest at the beginning of the academic term, can be useful to instructors and students in an Introduction to Psychology course. First, a sensitive pretest permits the instructor to evaluate the depth of preexisting
knowledge of course concepts, and the instructor can then adjust instruction to build on shared understanding or correct mistaken ideas. Second, by incorporating pretest questions into subsequent exams (in a pretest–posttest fashion), instructors can evaluate the extent to which course experience enhances student understanding. Third, by using individual differences in pretest results to predict course performance, an instructor can understand whether student achievement is influenced by preexisting differences in student knowledge and understanding. If the association is strong, the instructor may strive to assess sources of prior understanding and enlist these in the introduction of new information. Each of these goals concerns the value of pretesting to refining instruction or conducting outcome assessments of student learning.

Pretests can also benefit students more directly, especially when students understand that pretests are not measures of aptitude but rather of prior knowledge. A pretest previews ideas and concepts that students will encounter during the term, and pretests can also mobilize relevant prior knowledge about psychology. Moreover, when the pretest is similar in format to subsequent exams, it acquaints students with the instructor’s testing approach and style before they take tests that will contribute to the course grade. Students in our course have commented positively about their pretest experience for both of these reasons.

In this study, we administered a pretest on the second day of class in a large-enrollment Introduction to Psychology course. Throughout the semester, we subsequently included questions from the pretest in unit exams to assess improvements in student learning as a result of instruction. Although we expected significant gains in student understanding throughout the semester, we also anticipated that pretest scores would significantly and positively predict subsequent student performance on course exams. We based this expectation on several prior studies that have found significant correlations between pretest scores and course or exam grades in Introduction to Psychology, although none controlled for the influence of other variables that could also affect course performance (Carstens & Beck, 1986; Federici & Schuerger, 1976; Griggs & Jackson, 1988). The positive relation between pretest scores and course performance has not been consistent, however, perhaps because the diverse sources of prior knowledge of psychology can provide misleading as well as accurate understanding of psychological concepts. Thus it was also possible that pretest performance would have a negligible relation to subsequent student achievement in this course, as some have reported (Griggs & Jackson, 1988). Moreover, we were interested in determining whether pretest scores would be significantly predictive of overall course performance when we included other relevant predictors of student performance, including year in school, intended major, recitation performance, and other course activities.

Method

Participants

Students were 422 undergraduates enrolled in two sections of an Introduction to Psychology course at a large midwestern state university (ns ranged from 376 to 422 for specific analyses owing to attendance and enrollment fluctuations). The same instructor taught each section in an identical manner. Students from each section also participated together in weekly recitation sections. Consistent with typical enrollments for an introductory course in psychology, 40% of the students were men and 60% were women; 77% were first-year students, 15% were sophomores, and 4% each were juniors or seniors; 11% listed their intended major as psychology, with the remaining students undeclared (33%) or with intended majors in other fields. Most students were residents of the state and reflected the state’s predominantly White, middle income population.

Measures

The pretest consisted of a 25-item, five-alternative multiple-choice test. We limited the length of the pretest to 25 items (half the length of the standard course exams) to provide an appropriate assessment of prior knowledge without overwhelming students with a long exam on the second day of class. To create pretest questions, we surveyed psychology faculty members to identify the central concepts, issues, or ideas that they believed students in an introductory psychology course should know and sought to include the range of topics typically included in an introductory course. The pretest included questions about history and theories, research methods, brain and behavior, hereditary influences, states of consciousness, motivation and emotion, sensation and perception, classical and operant conditioning, memory, thinking and reasoning, developmental psychology, social psychology, psychological disorders, and personality theory and therapy. We selected topics for which formal or informal sources of prior knowledge in the popular media, folk theories, or everyday experience might be influential (e.g., interpersonal attraction, reinforcement, hemispheric specialization, sleep and dreaming). One question, for example, asked students to identify which of a series of folk sayings is confirmed by psychological research on interpersonal attraction, and the options included “familiarity breeds contempt,” “opposites attract,” “absence makes the heart grow fonder,” “birds of a feather flock together,” and “beauty is only in the eye of the beholder.”

Procedure

Our Introduction to Psychology course is a fairly conventional large-enrollment introductory course. Students meet twice weekly for a 75-min large-enrollment lecture section in a large auditorium, and once weekly in a 50-min small-enrollment recitation section led by a graduate teaching assistant. The lecture section consists of the presentation of new information through lecture with presentation software, discussions, and videos. The recitation section emphasizes demonstrations, informal experiments, discussions, and preparation for exams and other class assignments. Although there was no effort to record attendance at the lectures, teaching assistants consistently monitored student recitation attendance and the completion of weekly homework assignments and awarded points based on atten-
dance and homework that contributed to the overall course grade. Cumulative attendance and homework scores were thus dual indicators of the contribution of recitation to course achievement, along with student performance on a 25-item end-of-semester multiple-choice cumulative recitation exam.

After the pretest, students subsequently completed four unit exams at approximately 1-month intervals throughout the semester. Each noncumulative exam was identical in format to the pretest. Like the pretest, exam questions assessed students’ direct recall and comprehension of course concepts and their ability to apply these concepts to new situations and to integrate them in novel ways. Each exam incorporated six or seven questions that had previously appeared on the pretest (for five of the pretest questions, we made minor changes in wording for clarity).

Students also completed a five-page paper in which they applied course concepts to their analysis of one of four books chosen by the instructor (selections were The Crucible, Miller, 1953; Anne Frank: The Diary of a Young Girl, Frank, 1952/1993; Dibs in Search of Self, Axline, 1964; and Hamlet, Shakespeare, 1600/1963). Papers were due late in the semester, and the teaching staff graded papers using a 50-point scale, using criteria that included fulfilling content guidelines and writing mechanics. Therefore, student achievement in the course was indexed by four exam scores, the paper score, and the indicators of recitation performance described earlier. The analyses of student course performance thus included scores from four 50-item unit exams (each worth 50 points), the evaluation of a student paper worth 50 points, and scores awarded for attendance and homework assignments in recitation, together with scores on a 25-point recitation exam.

Results

Descriptive Analyses

Table 1 presents means and standard deviations for the pretest, the four exams, the paper assignment, and the recitation exam. Student performance on the exams was highly consistent throughout the semester, with mean scores ranging from 71% to 75% of the total possible points. Students’ performance was somewhat higher on the paper assignment and recitation exam (each at 80%). By contrast, the mean of 9.31 on the 25-item pretest was only 37% of the total points possible. The high score for the pretest was 19 (76%), by contrast with the perfect or near-perfect scores that were at the top of the range for the exams and paper assignment. Not surprisingly, therefore, students had some knowledge of psychological concepts on the second day of class, but their understanding was rather limited and incomplete.

Other psychometric properties of the pretest also contrasted with the four exams. Coefficient alpha, an index of the internal consistency of the test items, was .39 for the pretest, by contrast with the high alphas of .79 to .89 of the exams. The lower alpha of the pretest likely derived from its shorter length and the greater heterogeneity of the test items (by comparison with the exams). Taking these factors into account, however, the lower internal consistency of the pretest may also suggest that students’ understanding of one topic in psychology was not highly related to their understanding of other topics, which would be consistent with the informal, unsystematic sources of knowledge that probably contributed to pretest performance.

There was also, as expected, considerable variability in student performance on specific pretest questions, with the proportion of students answering questions correctly ranging from 8% to 66%. The pretest questions that the greatest proportion of students answered correctly concerned the characteristics of rapid eye movement (REM) sleep (66% answered correctly), the nature of schizophrenia (65%), the themes of developmental psychology (65%), and hemispheric specialization (55%). These topics are likely to be featured in news accounts or media feature stories, are relevant to personal interest or experience, and are commonly included in psychology courses in secondary schools (Carstens & Beck, 1986; White, Marcuella, & Oresick, 1979). By contrast, students performed most poorly on pretest questions concerning judgment heuristics (8%), the James–Lange theory of emotion (9%), and sympathetic nervous system (14%), each of which entails more specialized terms or knowledge specific to the field of psychology.

Student performance improved from the pretest to the course exams. The proportion of students who obtained the correct answer on pretest questions averaged 38% (range = 8% to 66%), by contrast with 77% (range = 28% to 91%) of the students who obtained the correct answer on the same questions when they subsequently encountered them on one of the four exams. On only six pretest questions did more than half the students answer correctly. By contrast, for the same questions included in the unit exams, more than half the students answered correctly all but one of these questions. On average, there was a gain of 39% in the proportion of students who answered each question correctly on the exam compared to the pretest. Interestingly, two of the three questions yielding the strongest performance (better than 90% answered correctly) when students responded to these questions subsequently on one of the four unit exams were the same as those eliciting the strongest pretest performance (i.e., REM sleep and schizophrenia), suggesting further the facilitating effects of prior knowledge.

Table 1. Descriptive Statistics for Student Performance Measures in Introduction to Psychology

<table>
<thead>
<tr>
<th>Measure</th>
<th>Range of Scores</th>
<th>Actual Range</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>0 to 25</td>
<td>3 to 19</td>
<td>9.31</td>
<td>2.79</td>
</tr>
<tr>
<td>Exam 1</td>
<td>0 to 50</td>
<td>19 to 49</td>
<td>37.60</td>
<td>5.95</td>
</tr>
<tr>
<td>Exam 2</td>
<td>0 to 50</td>
<td>10 to 49</td>
<td>35.58</td>
<td>7.16</td>
</tr>
<tr>
<td>Exam 3</td>
<td>0 to 50</td>
<td>15 to 50</td>
<td>35.93</td>
<td>7.78</td>
</tr>
<tr>
<td>Exam 4</td>
<td>0 to 50</td>
<td>10 to 49</td>
<td>36.62</td>
<td>7.70</td>
</tr>
<tr>
<td>Composite exam score</td>
<td>0 to 50</td>
<td>18 to 48</td>
<td>36.70</td>
<td>6.14</td>
</tr>
<tr>
<td>Paper assignment</td>
<td>0 to 50</td>
<td>17 to 50</td>
<td>40.27</td>
<td>5.16</td>
</tr>
<tr>
<td>Recitation exam</td>
<td>0 to 25</td>
<td>8 to 25</td>
<td>20.10</td>
<td>2.76</td>
</tr>
</tbody>
</table>
Interrelations Among Student Performance Measures

The intercorrelations among the student performance measures appear in Table 2 (N > 376). As expected, individual differences in exam performance were highly intercorrelated (mean $r = .68$), and exam scores were also positively related to scores on the recitation exam (mean $r = .49$) and the paper assignment (mean $r = .39$), even though the latter required somewhat different skills. By comparison, pretest scores were correlated with the other performance measures at a more modest but nevertheless significant level (mean $r = .34$; all correlations $p < .01$).

Predicting Student Course Achievement

Although pretest scores were significantly correlated with subsequent exam performance, it is also important to determine the unique contribution of pretest performance when considering other predictors of student performance. Consequently, the final analyses consisted of hierarchical linear regression models to predict student course achievement. Because of its significance to the overall course grade, the outcome measure we selected was the composite created by the mean of the four unit exam scores. Predictors were included in the following order: (a) two background variables (entered as a block): the student’s year in school (1 = first year; 4 = senior) and intended major (0 = nonpsychology; 1 = psychology), (b) pretest score, (c) paper assignment score, and (d) three indexes of recitation performance (entered as a block): score for cumulative attendance, score for homework assignments, and score on the recitation exam. The results appear in Table 3.

The addition of each set of predictors significantly incremented the proportion of variance explained in student overall exam performance. In the final equation, composite exam performance was significantly predicted by pretest scores, the paper assignment score, the recitation exam score, and the cumulative score for recitation homework. The background variables of year in school and intended major were not significant predictors in the final equation. Taken together, the regression accounted for nearly half (49%) of the explained variance in exam performance. Among these predictors, pretest exam performance was the second strongest, predicting 16% of the variance in exam scores.

We created similar regression models to predict each unit exam score based on student year in school and major (Step 1), pretest performance (Step 2), and scores for cumulative recitation attendance and homework prior to that exam (Step 3). We did not include scores for the paper assignment and the recitation exam in these models because they were primarily end-of-semester activities. The findings were consistent across regressions for each of the exams and were similar to the findings for the composite exam score reported in Table 3. In each, 26% to 28% of the explained variance in exam scores was predicted ($p < .01$). Among the significant predictors in the final equation, the pretest score accounted for 11% to 16% of the variance ($p < .01$) and was either the largest or second-largest (to recitation homework scores) in predicting exam scores.

### Table 2. Intercorrelations Among Student Performance Measures in Introduction to Psychology

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pretest</td>
<td></td>
<td>.41</td>
<td>.33</td>
<td>.35</td>
<td>.38</td>
<td>.42</td>
<td>.27</td>
<td>.28</td>
</tr>
<tr>
<td>2. Exam 1</td>
<td></td>
<td></td>
<td>.64</td>
<td>.63</td>
<td>.67</td>
<td>.82</td>
<td>.36</td>
<td>.47</td>
</tr>
<tr>
<td>3. Exam 2</td>
<td></td>
<td></td>
<td></td>
<td>.69</td>
<td>.71</td>
<td>.87</td>
<td>.36</td>
<td>.50</td>
</tr>
<tr>
<td>4. Exam 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.73</td>
<td>.89</td>
<td>.43</td>
<td>.45</td>
</tr>
<tr>
<td>5. Exam 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.90</td>
<td>.40</td>
<td>.54</td>
</tr>
<tr>
<td>6. Composite exam score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.45</td>
<td>.57</td>
</tr>
<tr>
<td>7. Paper assignment score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.31</td>
</tr>
<tr>
<td>8. Recitation exam score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** All correlations significant at or beyond $p < .01$, two-tailed.

### Table 3. Regression Analyses of Pretest and Other Predictors of Student Overall Exam Performance

<table>
<thead>
<tr>
<th>Step Variables Entered</th>
<th>$\Delta R^2$</th>
<th>$R^2$</th>
<th>df</th>
<th>$\Delta F$</th>
<th>$\beta$ (Step 1)</th>
<th>$\beta$ (Step 2)</th>
<th>$\beta$ (Step 3)</th>
<th>$\beta$ (Step 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Background</td>
<td>.03</td>
<td>.03</td>
<td>2, 360</td>
<td>4.94**</td>
<td>.13*</td>
<td>.10*</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td>Major Year in school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Prior knowledge</td>
<td>.16</td>
<td>.19</td>
<td>1, 359</td>
<td>73.00**</td>
<td>.12*</td>
<td>.06</td>
<td>.04</td>
<td>-.01</td>
</tr>
<tr>
<td>Pretest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Paper assignment score</td>
<td>.09</td>
<td>.28</td>
<td>1, 358</td>
<td>45.83**</td>
<td>.41**</td>
<td>.33**</td>
<td>.25**</td>
<td>.25**</td>
</tr>
<tr>
<td>Score for paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Recitation</td>
<td>.21</td>
<td>.49</td>
<td>3, 355</td>
<td>49.79**</td>
<td>.32**</td>
<td>.16**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative attendance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative homework score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation exam score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** All betas are standardized.

*p < .05. **p < .01.
Discussion

Students in Introduction to Psychology performed rather poorly on the pretest exam, but their performance improved significantly on the same questions that were included on subsequent unit exams, which suggests strengthened student understanding as the result of instruction. Nevertheless, individual differences in pretest performance were positively and significantly associated with every subsequent measure of course achievement. Moreover, the pretest accounted for a unique and significant proportion of variance in composite exam scores (and in unit exams) even when we included other predictors of student course achievement, such as paper and recitation performance, year in school, and intended major.

These findings suggest that the knowledge that students bring with them to the first day of class is positively and significantly predictive of their academic achievement in an introductory psychology course. Prior knowledge contributed to course achievement even though, relative to their subsequent achievement, students had relatively little understanding of psychological concepts as reflected in their pretest performance. These findings are consistent with broader literatures in psychology and education documenting the importance of preexisting understanding to new learning and the value of incorporating prior knowledge into instructional strategies. The large sample size of this study contributes to the strength and reliability of these results. Despite the diverse sources of prior knowledge in psychology, it appears that preexisting understanding facilitated, rather than undermined, student achievement in this introductory course.

Although these findings are important, it is noteworthy that the amount of variance explained in student achievement in Introduction to Psychology was nevertheless much lower than the level obtained in other studies of the effects of prior knowledge on later learning, where assessments of preexisting understanding explained 42% to 60% of the variance in follow-up assessments (Dochy, 1992; Tobias, 1994). The more limited contribution of prior knowledge in psychology may be due to the diverse sources of understanding available to students in psychology, which include formal coursework in secondary school, features in the popular media, folk wisdom, personal experience, and many other influences. Student achievement in an Introduction to Psychology course may build on earlier understanding derived from these sources, but it may also be undermined by misunderstanding and confusion as well. Moreover, this study is the only one to assess the influence of prior knowledge in psychology while controlling for other influences on student course achievement.

Fortunately, the findings of this research suggest that prior understanding has a significantly positive association with subsequent learning. The concepts with which students were most familiar on the second day of class were those on which they achieved greatest proficiency subsequently in the course, and the concepts students poorly understood related to specialized ideas in psychology and did not appear to derive from misconceptions from previous experience. Of course, a more systematic assessment of psychological knowledge is needed to confirm the conclusions of this study, but these findings suggest that instructors in introductory courses would be wise to build on students’ prior knowledge, rather than seeking to correct or ignore it, as a way of building student understanding of new terms and concepts. Building new instruction on students’ prior knowledge is also supported by research in cognitive psychology, especially from constructivist theory, which emphasizes the importance of enlisting prior understanding into the construction of new understanding (Committee on Developments in the Science of Learning, National Research Council, 1999). In an Introduction to Psychology course, therefore, it may often be appropriate to engage students in systematic reflection about what they already understand about a topic that the instructor will discuss. In our teaching, we increasingly begin the discussion of new topics by asking students “What do you know about …?” and build a lecture around what students have identified as elements of their prior knowledge. On other occasions, we use a news item or campus event to provoke student discussion of issues relevant to the next course topic, asking students to apply what they know of psychology to interpreting the event.

Do the effects of prior knowledge derive primarily from broader differences in student ability or aptitude? Unfortunately, we did not have access to admission test scores (e.g., the ACT or SAT) at the time this study to use as assessments of general student ability. Prior research has found, as one would expect, that SAT scores correlate significantly with measures of prior knowledge in psychology and with final course grades in introductory courses (Carstens & Beck, 1986; Griggs & Jackson, 1988). However, the amount of explained variance is low to moderate in each case, indicating that although general ability contributes to pretest performance and course achievement, other contributors (specific to developing understanding of psychology) are also influential. One important way of addressing this issue is, in future research, to add a measure of general ability to the regression analyses.

Despite the competencies students displayed at the beginning of the semester, students also improved significantly in their understanding as the semester progressed. This improvement was indicated by the stronger performance on pretest items when they appeared in subsequent unit exams (and, to a lesser extent, by the stronger internal consistency of the unit exams compared with the pretest). This improvement is not surprising after a semester of instruction in psychology, but it documents in ways that seldom occur the specific learning outcomes that can derive from students’ participation in an introductory course. In this regard, long-term follow-up assessments would be an additional, important way of evaluating the enduring benefits students derive from their class participation.

Although individual differences in pretest performance were associated positively with subsequent student achievement, it remains for future research to elucidate the determinants of pretest scores. Consistent with the educational literature, this discussion has emphasized prior knowledge derived from the many sources of information about psychology that people encounter in everyday life. However, there are other contributors to pretest performance, including students’ test-taking skills (e.g., question-attack abilities), vocabulary and verbal ability, and conscientiousness as well as others. Each of these qualities is likely to affect pretest performance and subsequent achievement on other course requirements, and they constitute a broader array of
personal resources (in addition to prior content knowledge) that students bring to the classroom on the first day. Further understanding of these qualities and their influence on academic achievement is important because it can potentially provide a multidimensional portrayal of the determinants of individual differences in learning in the collegiate classroom and the origins of these differences earlier in a student’s life experience.

References


Notes

1. We thank Dan Bernstein and Cal Garbin for thoughtful comments on an earlier edition of this article. We also thank Jennifer A. Haase, Sam Hardy, Sarah Kepple, Jennifer Kincaid, Haig Koumoudjian, Kristin Mollanen, and Sherill Pineda for their assistance with this research.

2. Send correspondence to Ross A. Thompson, Department of Psychology, University of Nebraska, Lincoln, NE 68588–0308; e-mail: rthompson1@unl.edu.
college courses. Ironically, one of the major goals of Levy and Peter’s research was to gather student data that would potentially afford faculty to align these faculty roles and student views.

Discussion

In sum, Bjornsen’s study contains value in that it expands on Levy and Peters by examining undergraduates’ open-ended responses regarding what they consider when evaluating college courses. Consistent with the stated goals of Levy and Peters, these data may serve to provide instructors with valuable insights for evaluating, revising, and improving their teaching and classes. In this regard the Bjornsen study serves as a useful corollary to Levy and Peters. However, contrary to its title, Bjornsen’s study fails to serve as a commentary to Levy and Peters or to expand and explain results of Levy and Peters. Bjornsen’s criticisms of the methodology, approach, and interpretation of results of Levy and Peters appear based on misperception of the scope and empirical goals of Levy and Peters. In addition, the many significant methodological and sampling difficulties in Bjornsen’s study make its direct application to Levy and Peters highly questionable.

Ultimately, successful pedagogy involves a complementary and mutual relationship between what students and faculty want and need, rather than a unilateral declaration by either individual party (McKeachie, 1999). Quite simply, any pedagogy that serves only student or faculty and fails to consider the unique institutional culture in which the pedagogy is engaged, is likely doomed to failure or at most mediocrity. Unfortunately, Bjornsen misperceives Levy and Peters as condoning a perspective where only desires and views of students and faculty are engaged, is likely doomed to failure or at most mediocrity. Unfortunately, Bjornsen misperceives Levy and Peters as condoning a perspective where only desires and views of students (or as he contends “consumers”) are served at the expense of pedagogical rigor and responsibility. In the end, a successful educational experience emerges when a balance between the wants and needs of both students and faculty is successful.

The unique institutional culture in which the pedagogy is engaged, is likely doomed to failure or at most mediocrity. Unfortunately, Bjornsen misperceives Levy and Peters as condoning a perspective where only desires and views of students (or as he contends “consumers”) are served at the expense of pedagogical rigor and responsibility. In the end, a successful educational experience emerges when a balance between the wants and needs of both students and faculty is present (Kennedy, 1997).

References


Note

Send correspondence to Gary Levy, University of Utah, Office of Budget and Institutional Analysis, 110 Park Building, 201 South President’s Circle, Salt Lake City, UT 84112; e-mail: gary.levy@utah.edu

Does Deliberate Source Monitoring Reduce Students’ Misconceptions About Psychology?

Joshua D. Landau
Anthony J. Bavaria
York College of Pennsylvania

Attempts to reduce students’ misconceptions about psychological principles by presenting contrary evidence have not successfully reduced the number of such misconceptions. The source-monitoring literature shows that carefully considering the source of memories reduces many memory-related errors. In this study, students judged the truth of common misconceptions about psychology. In a source condition, students had to report where they heard these misconceptions before making their judgment. This source manipulation did not reduce the number of misconceptions. Students in both conditions were more confident with their incorrect answers than with their correct ones.

Many students have misconceptions about psychological concepts (e.g., Brown, 1983; DeBell & Harless, 1992; Gutman, 1979; Lamal, 1979; Vaughan, 1977). For example, many students believe the myth that people use only about 10% of their potential brain power. Interestingly, introductory level psychology students (e.g., Gutman, 1979; Lamal, 1979), upper level psychology majors (Highbee & Clay, 1998; Lamal, 1995), nonmajors (Highbee & Clay, 1998), and the general public (Vaughan, 1977) all share these misconceptions.

Attempts to reduce the number of misconceptions have not been very successful (e.g., Lamal, 1995; McKeachie, 1960; Vaughan, 1977). Gutman (1979) found that exposure to accurate information during an introductory psychology course did little to correct people’s misconceptions about psychology. Thus, exposure to psychology course material is not a panacea for correcting inaccurate beliefs. However, there is evidence that instructing people to carefully scrutinize the source of their memories encourages them to carefully search memory for evidence pinpointing where they originally encountered particular facts (Johnson, Hashtroudi, & Lindsay, 1993). When people engage these extended retrieval processes, they are less likely to commit a multitude of memory errors (e.g., Landau, Thomas, Thelen, & Chang, 2002; Marsh, Landau, & Hicks, 1997; Multhaup, 1995).

To examine if these extended retrieval processes would reduce students’ misconceptions about psychological principles, people evaluated the truth of 10 inaccurate psychology statements. People in the source condition received instructions to carefully think about where they learned about each statement before making each judgment, whereas people in the no-source condition simply made their judgments.

Method

Participants

Ninety-two York College General Psychology students (Mage = 19) volunteered in this experiment in exchange for remote participation.

Participants were randomly assigned to one of two experimental conditions: the source condition and the no-source condition. In the source condition, participants were told that they would be presented with 10 statements that were commonly believed to be false, but for which evidence exists showing that they are actually true. They were also told that they would be asked to rate the truthfulness of each statement. Additionally, they were told that their judgments would be based on their memory of where they heard the statement, and that they should carefully think about where they heard the statement before making each judgment. In the no-source condition, participants were not given any instructions regarding the source of the statements.

Results

As predicted, people in the source condition were more likely to correctly identify the statements as false than people in the no-source condition. Specifically, people in the source condition were more likely to correctly identify the statements as false than people in the no-source condition, F(1, 90) = 25.76, p < .001. Furthermore, people in the source condition were more likely to identify the statements as false than people in the no-source condition, F(1, 90) = 25.76, p < .001.

Discussion

The results of this study support the idea that people are more likely to identify a statement as false if they carefully consider where they heard the statement. This finding is consistent with the source-monitoring literature, which suggests that carefully considering the source of memories reduces many memory-related errors. However, the results of this study are not consistent with the findings of previous studies that have used a similar methodology.

References

partial course credit and were randomly assigned to the source (n = 43) or no-source conditions (n = 49).

Materials and Procedure

Each participant received a booklet containing 10 false statements about psychological principles taken from Vaughan (1977) and Lamal (1979; see Table 1). Participants in the no-source condition read each statement and determined if it was true or false. After providing each answer, they rated their confidence using a 1 (not very confident) to 5 (very confident) scale. Participants in the source condition also read and judged each statement. However, before rendering their decisions, they had to record where they heard each fact, decide if the source was credible, and decide if it was reliable. After both groups responded to the psychology questions, they completed an 11-item demographic questionnaire by listing their gender, age, major, minor, current GPA, classification, psychology GPA, number of previous psychology courses, and if they had taken psychology in high school.

Results

Some participants did not complete the demographic questionnaire (n = 5) or failed to properly use the confidence scale; therefore, the degrees of freedom differ across the analyses. The students were primarily freshmen (72%) and sophomores (21%) who were taking their first college level psychology course (76%); many had taken a high school psychology course (43%).

Misconception Identification Accuracy

Even though people in the source condition had to describe where they encountered the information and decide if the source was credible and reliable, this instruction did not reduce their propensity to claim that the majority of these false psychological statements were true. People in both the source (30%) and no-source conditions (29%) accurately identified the same number of misconceptions, \( F(1, 90) = .04, p > .05 \).

Confidence Ratings

We calculated each participant’s mean confidence ratings for their correct answers (i.e., called false) and compared them with their confidence rating for their incorrect answers (i.e., called true). The confidence ratings for those in the no-source condition were 3.32 (SD = .62) and 3.71 (SD = .43) for their correct and incorrect answers, respectively. People in the source condition provided similar confidence ratings for their correct (M = 3.14, SD = .03) and incorrect answers (M = 3.83, SD = .92). A 2 (condition: source versus no source) × 2 (response type: incorrect versus correct) mixed ANOVA yielded a significant main effect for response type, \( F(1, 85) = 26.3, p < .05 \), which indicated that people in both conditions were more confident with their incorrect answers than with their correct ones. Neither the main effect of condition nor the Condition Type × Condition Interaction were significant, both Fs < 2.03.

To investigate how students responded to each item, we collapsed across the experimental conditions and calculated the percentage correct and the mean confidence ratings for correct and incorrect answers (see Table 1). This analysis indicated that accuracy ranged from 12% correct for the blind and minority statements to 55% correct on the gender statement. Comparison of the correct and incorrect confidence ratings indicated that for 8 of the 10 statements, confidence ratings were higher when people incorrectly believed that the statements were true. There were only 2 statements for which people displayed higher confidence when they were correct. This pattern of results clearly demonstrates that these confidence effects are not restricted to just a few of the statements.

Sources of the Misconceptions

People in the source condition described where they originally encountered each statement, which allowed us to ex-

| Table 1. Percentage Correct and Correct and Incorrect Confidence Ratings for Each Statement |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Item                            | % Correct | Correct Confidence | Incorrect Confidence |
| The basis of the baby's love for his mother is the fact that his mother fills his physiological need for food. | 32        | 3.72                | 3.32**                        |
| Most humans use only 10% of their potential brain power. | 23        | 2.90                | 3.51*                         |
| Blind people have unusually sensitive organs of touch. | 12        | 3.18                | 4.05*                         |
| The more motivated you are, the better you will do at solving a complex problem. | 20        | 3.33                | 3.97*                         |
| The weight of evidence suggests that the major factor in forgetting is the decay of memory traces with time. | 29        | 2.63                | 3.52*                         |
| Psychotherapy has its greatest success in the treatment of psychotic patients who have lost touch with reality. | 51        | 2.72                | 3.18                          |
| Personality tests reveal one's basic motives including those you may not be aware of. | 30        | 3.00                | 3.66**                        |
| To change people's behavior toward members of ethnic minority groups, we must first change their attitudes. | 12        | 3.36                | 4.10*                         |
| The study of the mind is the best brief definition of psychology today. | 33        | 4.03                | 3.94                          |
| Boys and girls exhibit no behavioral differences until environmental influences begin to produce such differences. | 55        | 3.67                | 3.95                          |

Note. We compared people’s confidence levels for each question using between-subjects t tests. The confidence scale ranged from 1 (not very confident) to 5 (very confident).

*p < .05. **p < .09.
amine the different sources and the percentage of correct and incorrect responses associated with each source (see Table 2). Across the 430 responses, there were 14 cases where people left the source question blank, leaving 416 classifiable responses. Participants described 31 different sources that we condensed into six categories. The psychology-related category included the psychology instructor, psychology lecture material, psychology classroom discussions, psychology handouts, or psychology textbooks. The other course category included college and high school courses outside of psychology (e.g., math class, seminars, conferences). The vague source category reflects situations in which people told us that “everyone knows” a fact or when they attributed a statement to common sense or other forms of shared knowledge. The media category refers to TV, magazines, the Internet, or movies/videos. When participants attributed a statement to friends, family, or coaches, we counted these in the personal contact category. The never heard before category reflected instances where participants told us that they had not heard the statement before.

Inspection of Table 2 reveals several interesting results. First, when people said that they had not heard a statement before, they were more likely to claim that it was false (40%) rather than true (11%). This finding indicates that participants were not simply calling statements true. If they could not think of where they heard a statement, then they adopted a more conservative response criterion for verifying a statement. Second, when students claimed that a misconception was true, they tended to attribute this knowledge to the psychology-related source or other courses (58.4%). Finally, participants reported that the media, common sense, and their friends and families were equally responsible for correct and incorrect judgments.

Discussion

Even though we required students to describe where they heard about false statements about psychology and to judge the credibility and reliability of the source, they were just as likely as the control group to claim that the majority of the misconceptions were true. Interestingly, people were less confident when they claimed that a statement was false compared to when they claimed that a statement was true. This result provides some insight into the cognitive processes people engage when judging these statements. Many of these misconceptions are widely held in our society and are repeated by a wide variety of sources. Therefore, when people question their truthfulness and correctly mark the statement as false, this choice raises some doubts about their answer and they compensate by providing a lower confidence rating.

Perhaps the more discouraging aspect of these findings is that participants claimed that they encountered these misconceptions about psychology in educational settings (psychology or other classes). There are at least three potential reasons why students attributed their misconceptions about psychology to this source. First, it is possible that their instructors presented the information in a way that led the students to believe these misconceptions were actual psychological facts. Second, because the statements referred to psychological concepts, students might have simply assumed that their instructor or the text was the source of the information. Third, it is possible that students were confused about the information in some of the statements. For example, students may remember covering information about Harlow’s (1958) research, but they might not remember the exact details of the study. Therefore, when judging this statement, they forgot that compared to food, warmth and physical contact were better explanations for love. In this case, students simply misremembered the details that are key to correctly judging each statement.

In conclusion, our findings demonstrate that requiring people to carefully consider where they learned about these inaccurate statements is not an effective way to reduce people’s belief in their misconceptions about psychology. Although manipulating the degree to which people engaged source-monitoring processes did not reduce their erroneous beliefs about psychological principles, these results are of some value because they suggest that this technique is not effective for reducing these tenacious myths about psychology. Clearly, more empirical work is needed to develop effective techniques that will reduce the number of psychological misconceptions.

Table 2. Percentage of Responses Associated With Each Source in the Source Condition

<table>
<thead>
<tr>
<th>Source</th>
<th>Correct (&quot;False&quot;)</th>
<th>Incorrect (&quot;True&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology-related information (text, instructor, handout)</td>
<td>24.6</td>
<td>37.9</td>
</tr>
<tr>
<td>Other courses (sociology, health, high school)</td>
<td>5.1</td>
<td>20.5</td>
</tr>
<tr>
<td>Vague source (common sense)</td>
<td>11.9</td>
<td>12.4</td>
</tr>
<tr>
<td>Media (TV, magazines, Internet, etc.)</td>
<td>9.3</td>
<td>11.1</td>
</tr>
<tr>
<td>Personal contact (family, friends, etc.)</td>
<td>8.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Never heard before</td>
<td>40.7</td>
<td>11.7</td>
</tr>
</tbody>
</table>

References


Notes

1. We thank Randolph A. Smith, Andy Leynes, and three anonymous reviewers for their helpful comments on an earlier draft of this manuscript.

2. Send correspondence to Joshua D. Landau, Department of Behavioral Sciences, York College of Pennsylvania, York, PA 17405–7199; e-mail: jlandau@ycp.edu.

A Motivating Exercise for the Introductory Class (and Beyond)

Louise Katz
Columbia State Community College

This article describes an exercise used with 12 community college psychology classes that produces high levels of motivation and offers the opportunity for a personally meaningful experience. Students choose a topic of personal interest related to psychology and investigate it through student-conceived field experiences and literature research. The exercise includes a poster presentation, hand-out, and oral presentation or, alternatively, a term paper and journal. At least 78% of 181 students reported selecting a topic they hoped would be useful in their personal life, and 88% reported it had been. Many reported being helped with career decisions. In classes with oral presentations, students learned from each other’s research, and 93% reported finding listening to classmates’ presentations worthwhile.

Instructors often struggle to develop strategies to motivate students. In this article I describe an exercise that I have found very motivating with community college students, that can be carried out even with limited academic resources and may be used or modified for use with a variety of psychology courses in 2- and 4-year institutions. The exercise provides an academically and personally meaningful experience, greatly engages student interest in psychological topics, and creates appreciation of the value of psychological research through “hands-on” real-life experience. Brems (1994) reported that early familiarity with the research process may facilitate lessened anxiety in upper level courses. Others have discussed the value of journal writing (Connor-Greene, 1990; McCluskey-Fawcett & Green, 1992; Perry, Huss, McAuliff, & Galas, 1996; Sugar & Livosky, 1988). However, there is an absence in the literature of projects for freshmen where students both select their topic to investigate and fashion a means to explore it in the “real world.”

Method

At a rural community college, 199 predominantly freshman students in 10 introductory psychology classes and 2 life span development survey classes participated in the exercise. Students in 5 classes presented their work by submission of a journal and term paper, and students in 7 classes presented their work with oral presentation to the class; I discuss these two approaches separately. Approximately 73% of the students were women and 27% were men; some were returning adult students. Students chose a topic of personal interest related to psychology and prepared a one-page proposal, due 3 weeks after the beginning of the semester, that included a minimum of two proposed field experiences conceived by the student to investigate the topic and the proposed area or areas of reading related to the topic. Prior to the proposal due date, I discussed types of possible field experiences with the class, such as observations, interviews, attendance at meetings or support groups, surveys, and simple experiential projects, and I explained interview techniques, confidentiality, and securing of appropriate permissions. Also prior to the proposal due date, the class attended a presentation by the librarian on methods of library and database research, and students shared potential projects in a class discussion and received feedback from me. I returned written feedback on the proposals or, with some classes, met individually with each student during a class session to give oral feedback. Review of proposals gives the instructor the chance to ensure that projects do not pose ethical, student welfare, or institutional liability concerns. I reviewed proposals I had concerns about with another licensed psychologist knowledgeable in ethical matters, and I redirected these students to a related emphasis in their field experience, typically an interview. A few weeks later, I invited students to raise questions in class and offered them the opportunity to discuss their progress with me individually. I required five research sources for the exercise; I permitted two of these to be Web sites. The project as a whole represented 33% of the final course grade.

Presentation Approach

The presentation approach, used in seven classes, required students to present their work with an oral presentation, poster, and handout. Each student created a poster for a poster session held during the class session before oral presentations began. I explained possible approaches to poster organization in class and encouraged creativity. I advised students to include in their posters information gained from both their literature research and field experiences. Oral presentations took place over a period of one or more weeks near the end of the semester. Some students expressed anxiety about presenting a speech. It was helpful to discuss how to prepare for and present a talk. I made available a pamphlet on public speaking anxiety (Katz, 2000). Each student prepared a handout that covered the main points learned from literature research plus a list of references used and furnished a copy to each classmate and the instructor at the time of the oral presentation. I typically required a presentation of 20 to 25 min, followed by the opportunity for questions and comments. I found that a length requirement of at least 15 min encouraged students to prepare thoroughly. To permit alternatives for anxious students, I gave
Empowering Students: Class-Generated Course Rules

Jeannie D. DiClementi 
Indiana University–Purdue University Fort Wayne

Mitchell M. Handelsman 
University of Colorado at Denver

After we gave 2 classes of introductory psychology students the syllabus, the first class (the experimental group) generated rules for classroom behavior. The instructor presented the second class (the comparison group) with the list of rules and said they were instructor generated. Students rated the rules, several aspects of the course, and the instructor. The comparison group (n = 88) reported higher frequencies of negative behavior by class members. Students in the experimental class (n = 62) rated the instructor more positively. The groups did not differ in grades, perceived fairness, or perceived importance of the rules.

Instructors often complain when students appear not to be engaged or have negative attitudes toward their courses (Sacks, 1996). We believe the syllabus can be helpful in attenuating students’ attitudes. Many have written about what information to put into a syllabus (e.g., Appleby, 1999; Davis, 1993), but less about how instructors can use the processes surrounding the syllabus, especially during the first class meeting (Perlman & McCann, 1999), to create an optimal learning environment.

In addition to obtaining important course information, students will “read between the lines” of the syllabus to glean other critical information about such instructor characteristics as interpersonal style, attention to details, availability, and approachability. Students’ perceptions of the course and of the instructor may be influenced by this implicit information, and their perceptions could determine their level of engagement in the course. For example, if the syllabus is extremely formal, organized, and structured, students may perceive inflexibility and rigidity and be less inclined to go to
the instructor for help. A laissez-faire approach may give students the message that the course material or associated requirements are not important. Consequently, students may spend less time studying or may put less effort into required projects.

Unfortunately, students may get a wrong or incomplete impression about a course for at least two reasons. First, students may not be good at attending to the messages sent, either explicitly (Becker & Calhoon, 1999) or implicitly, by the syllabus. Second, instructors may not be aware that the messages their syllabi send create impressions that are inconsistent with their teaching philosophies or styles.

Although rules for student behavior are important, neither students nor faculty may attend to them sufficiently. In addition to policies on cheating and plagiarism, such rules include expectations for cell phones and pagers, eating in class, addressing the instructor, and leaving early. Many instructors have found it easier to list the rules on the syllabus or in an attachment than to correct student behavior in the middle of a lecture.

A side effect of having lists of rules may be that they implicitly communicate an adversarial relationship between instructors and students. When students discover that instructors expect them to behave irresponsibly, their reactions might be negative. Such a syllabus may create a self-fulfilling prophecy: Students may test the limits, and instructors may feel justified as well as encouraged to make even more rules. The result can be a disaster for both the faculty member and the students.

Given that some form of guidelines or rules for the class is necessary to socialize students into the course and that communicating the rules can have serious implications for engagement, how can instructors create expectations for appropriate behavior and still maintain an atmosphere of respect and cooperation? Clinicians talk about empowering clients to make their own decisions and to take responsibility for their actions. Instructors can adapt this approach in the classroom to empower students to become responsible classroom citizens. In this article we (a) present a method for empowering students and (b) assess whether empowering students to be responsible for their classroom behavior had any effect on such behavior, on their perceptions of the course and of the instructor, and on their grades.

The first author implemented a strategy intended to empower her students in an introductory psychology course. Students in one class (the experimental group) generated rules for classroom behavior. The second class (the comparison group), received the rules generated by the first class, but believed the rules to be instructor generated. We predicted that the students choosing their rules would assess the course experience more favorably in terms of not only the rules, but also the class in general and the instructor.

Participants in the study were students enrolled in two introductory psychology classes taught by the first author. These classes met at 8:00 a.m. and 9:00 a.m. on a 3-day-a-week schedule. On the first day of class, prior to receiving the syllabus, students in each class completed a brief questionnaire that asked sociodemographic information and questions related to taking the class: “What grade do you expect to get in this course?”; “How would you rate your knowledge of psychology right now?” (1 [not at all], 4 [about average], 7 [very knowledgeable]); and “How would you rate your interest in the field of psychology?” (1 [very low], 4 [about average], 7 [very high]). Finally, they rated the importance of 21 items normally found on a typical course syllabus, each on a scale ranging from 1 (not at all important) to 7 (very important)

In each class, after completing the questionnaire, students then received the course syllabus and spent the next several minutes discussing the information detailed in the syllabus. The 8:00 class then divided into small groups of 5 students each. Each group received a rule category listed on the last page of the syllabus and developed a rule for that category. Categories included eating in class, sleeping in class, coming in late, and use of phones and pagers. The entire class voted on each of the suggested rules and agreed on strategies for self-management of rule violations. Each student wrote the rules in the allotted spaces on the syllabus.

Surprisingly, some of the rules were stricter than those the instructor might have chosen. The class was able to compromise with relative ease, however. The only concrete suggestions they requested were the options for addressing the instructor. The instructor listed these as title (“Dr.”) and last name, first name, or “Professor” and last name. Creatively, the students chose “Dr. D,” which had not been one of the options.

When the 9:00 class (comparison group) reached the last page of the syllabus, the instructor explained to the students that to facilitate recall of the rules, she was going to present the rules orally while the students wrote them in their syllabus. The instructor then read the list of rules generated by the earlier class. That way, both classes had the same set of rules and an identical syllabus. The difference was that the instructor told the 9:00 students that these were her rules.

At the midpoint and again at the end of the semester, students completed several dependent measures (all on 7-point Likert scales). They rated the importance of the rules, the rules as a learning experience, and the fairness of the rules. They also estimated the frequency of nine classroom behaviors (e.g., “Classroom noise level too high,” “Students being rude to other students”) on a scale ranging from 1 (none at all), 4 (average amount), to 7 (an extreme amount). Finally, they rated their perceptions of the instructor.

We conducted this study on the regional campus of a large midwestern university. This is a commuter campus enrolling approximately 12,000 students, of which about 54% are full-time. The average age of students on this campus is 26 years.

**Method**

Demographics

A total of 150 students, 62 (out of 67 initially enrolled) in the 8:00 class and 88 (out of 95 possible) in the 9:00 class, completed questionnaires at the beginning of the semester,
and we collected data at Week 8 (ns = 44 and 60) and at Week 15 (ns = 35 and 49). Fifty-six percent of each group completed the questionnaires at Week 15, so we did not have differential drop or attendance rates.

The average age of the students was 20.72. Women constituted 59.3% of the group (n = 89); men, 40.7% (n = 61). One hundred twenty-five (83.3%) of the respondents listed themselves as White, 16 (10.7%) as African American, and fewer than 5 each as Hispanic, Asian/Pacific Islander, or multiracial. Most of the respondents were first-year students (n = 126; 84.0%), 17 (11.3%) were sophomores, 4 (2.7%) were juniors, and only 1 student (0.7%) was a senior. Two participants did not list their class. The participants in the two groups did not differ in terms of age, grade-point average, expected grade, reported knowledge of psychology, or reported interest in psychology.

Importance of Behavioral Rules

Participants at Week 1 rated the importance of items on the syllabus. One of these items was “Rules for behavior in the classroom.” An ANOVA on this item with the between-subject factors of group and sex yielded a main effect for sex, F(1, 146) = 5.26, p = .023, such that women saw the rules as more important (M = 5.17) than did men (M = 4.64). There were no effects for group and no significant Group × Sex interaction.

Fairness of Rules

At Weeks 8 and 15, students rated the fairness of the classroom behavior rules. A MANOVA yielded no significant effects.

Learning Experiences

Participants rated several aspects of the course (e.g., videos, lecture, research participation) for the value as learning experiences. Three of these items were relevant to this study: rules for classroom behavior, syllabus, and class discussion. A MANOVA with these three dependent variables yielded a significant Group × Sex interaction, F(6, 69) = 3.00, p = .012. The only univariate ANOVA that was significant was for rules for classroom behavior at Week 15, F(1, 74) = 7.81, p = .007. Subsequent tests revealed that women in the experimental group rated these rules as less important (M = 3.42) than did women in the comparison group (M = 4.58), F(1, 50) = 6.06, p = .017. Men did not differ in their ratings.

Frequencies of Observed Behaviors

At Weeks 8 and 15, participants estimated the frequency with which nine behaviors occurred in class. Six of these behaviors were negative; three were positive. We performed a MANOVA on the six negative behaviors, both at Weeks 8 and 15. We did not use time as an independent factor because we were not interested in changes over time, only in seeing if there were effects at any time during the semester.

The MANOVA yielded a main effect for group, F(12, 62) = 3.71, p < .001. Subsequent univariate ANOVAs revealed that participants in the comparison group reported higher frequencies of negative behaviors than those in the experimental group at both times, with the exception of students leaving early, which reached significance only at Week 15. The other behaviors that reached significance at both times were classroom noise level too high, students coming in late, students being rude to other students, students being rude to instructor, and students talking to each other during class. A MANOVA performed on the three positive behaviors (good student participation in class discussions, good cooperation among the students, and students enforcing the rules with each other) yielded no significant differences.

Instructor Ratings

Participants answered nine questions, at Weeks 8 and 15, regarding qualities of the instructor. A MANOVA on these ratings yielded a main effect for group, F(18, 57) = 2.07, p = .019. Univariate tests showed that students in the experimental group rated the instructor as more courteous toward students at both Weeks 8 and 15 than did students in the comparison group. At Week 15, ratings were also higher in the experimental group on the following items: willingness to answer questions in class, willingness to hear different points of view, encouragement of classroom discussion, and genuinely interested in students. In summary, ratings of the instructor were higher when discussion about the rules occurred. The comparison group did not rate the instructor higher on any items.

Course Grades

There was no significant difference in course grades between the groups.

Discussion

These results provide some support for the effectiveness of our procedure. According to the perceptions of students, fewer violations of the rules occurred when students had the opportunity to develop their rules. Whether these behaviors were actually occurring less frequently, student perception of the classroom environment may have an important impact on the students’ classroom experiences. Feeling comfortable in an environment and having a sense of control over their experiences may enhance students’ investment in the class.

An unexpected observation came from the instructor’s department chair during the semester. As part of the department evaluation process, the chair observed the instructor’s classes once during each semester. The chair commented later that the second class was noisier with more students talking among themselves during the lecture.

Students in the empowered class also had more favorable attitudes toward the instructor. They rated the instructor as more courteous, more willing to answer questions and to hear different points of view, and more encouraging of classroom.
discussion. These perceptions may have stemmed at least partially from the first day of class when the instructor asked the class to generate its set of rules. The instructor's announcement may have signaled to students that they were going to have a different experience, and students may have altered their perceptions of the instructor throughout the semester. Time spent on the rules was different for each class, and this difference may have confounded the results. Limitations of these data include the difference in class sizes and times of day the classes met. A further limitation of these data is that the instructor was not blind to the experimental condition, which could have affected her behavior in both classes. However, an unexpected benefit of empowering the students could be a corresponding change in the instructor's behavior that then results in a more positive classroom experience.

The positive impact of this procedure was limited in this study: The procedure did not influence students' grades, their perceptions of the fairness of the rules, or their rating of the importance of the rules. Indeed, among women, those in the experimental class rated the classroom rules as less important than did those in the comparison class. They may have been responding to the length of time spent discussing the rules and may have perceived that the rules did not warrant extra discussion.

Instructors could adapt this procedure to meet the needs of a wide variety of classroom situations. Classes could discuss fewer categories of rules, giving them the chance to discuss them at greater length, particularly those that have been most problematic for instructors. Instructors may wish to present students with finite lists of options for rules to avoid students developing rules that the instructor cannot live with. Soliciting feedback from students about the procedure and the rules could not only add to the students' sense of empowerment, but could help the instructor to fine-tune the process in future classes.

This one study is not a definitive test, but it appears that the procedure at least does no harm and may be useful. At the very least, discussion of classroom rules may serve to open communication between instructor and students. Students may feel more comfortable approaching the instructor with questions or concerns, which would be an additional benefit.

References


Note

Send correspondence to Jeannie D. DiClementi, Department of Psychology, Indiana University–Purdue University Fort Wayne, Neff Hall Room 388D, 2101 East Coliseum Boulevard, Fort Wayne, IN 46805; e-mail: diclemef@IPFW.edu.
Student Use of Introductory Texts: Comparative Survey Findings From Two Universities

Jason F. Sikorski
Kelly Rich
Bryan K. Saville
William Buskist
Auburn University

Oksana Drogan
Stephen F. Davis
Emporia State University

We surveyed introductory psychology students at 2 universities regarding their purchase and use of introductory level college texts. Most students who purchased texts used them infrequently, perceived that studying class notes and attending lectures were more important than reading the text for receiving a good grade, and spent less than 3 hr per week reading their texts. We recommend that college instructors establish specific, predictable contingencies to increase the likelihood that students will purchase and read the assigned texts in their introductory classes.

Reading is central to the pursuit of individual learning in college and university settings (Steuer, 1996). Thus, many college and university courses are typically organized around a textbook. Because the text plays such a prominent role in college courses, several psychologists study text characteristics and their relation to specific academic disciplines. Indeed, we now know much about introductory psychology texts (e.g., Fernald, 1989; Griggs, Jackson, Christopher, & Marek, 1999; Weiten, 1988; Zeckmeister & Zeckmeister, 2000) and the publishers who develop and produce them (e.g., CUSH & Buskist, 1997; Griggs et al., 1999).

We know much less, however, about how students use introductory texts in their coursework. Students rate bold-faced words, chapter summaries, running glossaries, and self-tests as helpful learning aids (Weiten, Deguara, Rehmke, & Sewell, 1999; Weiten, Guadagno, & Beck, 1996). Gathering additional information about how students use their texts would seem useful in developing strategies for enhancing student learning in introductory psychology and other introductory courses. Toward this end, we compared introductory psychology students at a large, Southeastern institution (Auburn University [AU]; Doctoral/Research Extensive) with students at a smaller Midwestern institution (Emporia State University [ESU]; Masters’ College and Universities I) in their use of texts for introductory college classes including introductory psychology.

Method

Participants

We administered a survey to 439 AU students (217 women and 222 men) and 739 ESU students (464 women and 275 men). The majority of students from both universities were either freshmen or sophomores (82% at both schools). Students from the AU sample were enrolled in one of three sections of an Introductory Psychology course taught by one instructor during the middle of the Spring semester, 2000. Students from the ESU sample were enrolled in one of eight sections of an Introductory Psychology course taught by eight different instructors during the middle of the Fall semester, 2000. The textbooks utilized in the introductory courses at both universities were described in the syllabus as being required reading. Chi-square analyses revealed no significant sex differences or differences between freshmen and sophomores relative to juniors and seniors for any of the comparisons described subsequently.

Procedure and Survey Instrument

After reading an informed consent sheet, all participants completed an 11-item questionnaire. The first 2 items queried students regarding their year in school and sex. Three of the survey questions dealt specifically with introductory psychology courses; the remaining 6 questions concerned introductory courses across all college disciplines.

Results and Discussion

Thirty one percent of AU participants reported purchasing the introductory psychology text compared to 91% of ESU students. This difference is likely due to the fact that the AU students were using a book that the local bookstores would not buy back. Thus, most AU students who did not purchase the text (76%) reported borrowing it from a friend, as an ample supply of the books was available free of charge. The ESU students used a textbook that their bookstore was buying back. Fifty-six percent of ESU students who did not purchase a text reported borrowing it from a friend.

Thirty percent of AU students responded that they did not purchase an assigned text for at least one other introductory college course; 18% of the ESU students reported the same. Again, the primary reason cited by AU students (32%) is that they borrowed the text from a friend. Other AU students reported that they did not purchase a text because of financial concerns (13%), they scored better on exams by using their
notes only (10%), or they thought that the text was "useless" or that they would not need it (18%). Twenty-five percent of the AU students provided other reasons (e.g., the teacher lectured from a different book, the course was a review, or the student was "too lazy"). Among the reasons cited by ESU students for not purchasing a text for at least one other introductory college course were that they believed the book was not necessary (54%), too expensive (21%), or they borrowed it from a friend (17%). Eight percent of the ESU students reported that they just never got around to buying it.

Although most students reported buying texts for their various introductory courses, they did not necessarily always read them. Eighty-two percent of the AU students and 78% of the ESU students reported either not reading a text for at least one introductory college course or reading it sparingly after purchasing it. However, only a small minority at both institutions reported that they believed reading their texts was "not at all important" or "not very important" in determining their final grade in their introductory courses (AU = 11%; ESU = 21%).

Nonetheless, when asked to choose the single most important factor for doing well in introductory courses, most students from both institutions chose other factors. Both the majority of AU students (64%) and ESU students (58%) reported that taking notes and studying them (without reading the text) was the single most important contributor to doing well. Students from both institutions (AU = 29%; ESU = 37%) also rated attending class and listening to the lecture as the single most important factor more often than only reading the assigned text for the typical introductory college course. Only a small fraction of the students from both institutions (AU = 6%; ESU = 4%) reported that only reading the text was the most important factor influencing their final grades. These findings, however, should be considered in the context of the forced-choice nature of the related survey item. Obviously, all of the factors rated are likely very important, but we instructed students to choose only the most important factor.

Students from both institutions also reported not reading their texts very much for typical introductory college courses. For a given week, the majority of students from both institutions reported spending less than 3 hr reading their texts (AU = 80%; ESU = 93%). In addition, the majority of students from both schools reported that they would not begin reading for an examination covering two chapters in their introductory psychology course scheduled in 1 week until 3 days prior to the examination (AU = 60%; ESU = 70%). Fewer than 20% of the students from both institutions reported that they would begin reading 1 week or more in advance (AU = 18%; ESU = 14%).

Although the results are somewhat qualified because of the possible response biases created by the forced choice and restricted range nature of some of the survey items, we believe that purchasing and reading an introductory text plays only a minor role in how students tend to prepare for examinations in introductory level courses across disciplines in college. That AU and ESU students showed similar patterns in their use of introductory texts across different courses suggests that problems associated with the minimal reading of texts may not be limited to geographic region or institutional size.

These results highlight the importance of teachers structuring their courses in ways that promote frequent and critical reading of the textbook. Providing students with specific and predictable consequences for making the choice of whether to read the textbook may alter the frequency and regularity with which students use their textbooks in introductory courses across college disciplines.

Quizzing students over reading material and calling on students in class to discuss text material are both effective ways of getting students "into" their textbooks well in advance of examinations (Davis, 1993; McKeachie, 1999). In addition, student reading of the textbook is probably more likely when instructors' class presentations highlight textual information without explicitly duplicating it. After the first examination, students are able to ascertain whether it is necessary to read the textbook to reach their academic goals in the course. Examinations that do not contain material from the textbook are not likely to promote student reading and studying. Students who succeed in introductory level college courses without critically reading their texts would seem less likely to excel in future individual learning endeavors in college and beyond.

References


Notes

1. We thank members of the EDGE research group on teaching for reading and commenting on several earlier versions of this article.
2. Send correspondence and requests for copies of the survey used in this study to Jason F. Sikorski, Psychology Department, Auburn University, Auburn, AL 36849-5214; e-mail: sikorf@auburn.edu.
view as an externally imposed requirement that all researchers must face. Clearly, conveying a balanced message makes more sense than conveying a negative view of the IRB as a common enemy or as creating “unwarranted meddling” in one’s research. Obviously, the message to students should not be that it is acceptable to ignore IRB procedures when they are under time pressure. Rather, instructors should point out to students that faculty can feel similar pressures when they conduct their research. The information I present to my students provides a good illustration of those pressures felt by faculty researchers. I have collected data showing that students who received and discussed this information show attitudes about ethical issues that are closer to faculty attitudes than students who did not receive it.

In conclusion, the major value of this approach is that it focuses directly on a neglected aspect of teaching and learning research ethics: the nature of the researcher–IRB relationship and how it can affect the research process at various stages. By focusing on this relationship, instructors have an additional tool for enhancing the ethical understanding of students. Even if students normally gain some understanding of the specific do’s and don’ts of ethical psychological research, incorporating this type of information may provide them with a more accurate appreciation of the bigger picture that researchers face. This information can be useful in beginning to complete that picture for students and in giving them a better appreciation of the dynamic nature of the researcher–society social contract Rosnow (1997) described.

References


Notes

1. Preparation of this article was supported by a Middle Tennessee State University Non-Instructional Assignment grant.
2. I thank Will Langston, John Pennington, Terry Whiteside, and the editor and reviewers for their comments on earlier drafts of the article.
3. Send correspondence and requests for further details on the faculty respondents, survey instruments, procedure, and evaluation data to Tom Brinthaupt, P.O. Box X034, Department of Psychology, Middle Tennessee State University, Murfreesboro, TN 37132; e-mail: tbrinthua@mtsu.edu.

Textbook Selection: Balance Between the Pedagogy, the Publisher, and the Student

R. Eric Landrum
LuAnne Hormel
Boise State University

We report a national survey of introductory psychology instructors about how pedagogical aids and publisher practices impact textbook selection and student learning. Respondents evaluated the importance of 79 attributes on 2 separate criteria: textbook selection and facilitating student learning. Some criteria received high ratings on both dimensions (e.g., textbook accuracy, readability/writing quality, use of examples on both textbook selection and student learning), whereas others received high ratings on only 1 dimension (i.e., ancillary materials and test item files for the textbook selection process). Experience in teaching the course also influences the relative importance of textbook selection criteria.

“One of the most important decisions an instructor makes is the selection of a textbook” (Chatman & Goetz, 1985, p. 150). Although not all instructors would agree with the previous statement, this decision is both an important one and also a difficult and daunting task. Griggs, Jackson, Christophe, and Marek (1999) reported that there are 37 full-length introductory psychology textbooks available; there are numerous other “basic” or “briefer” editions available. For instructors a critical question becomes “Which textbook do I choose?” Our interest was to understand (a) the process of selecting a textbook and (b) the instructor’s perceptions of textbook’s role in facilitating student learning. Most of the literature on textbook selection is advisory in nature. These suggestions offered different methods used (Chatman & Goetz, 1985), including a textbook selection committee (Britton, Woodward, & Binkley, 1993), student
involvement in the process (Griesinger & Klene, 1984), and focusing on the reading abilities of students when selecting a textbook (Stephens, Weaver, Ross, & Emond, 1986). When students have evaluated textbooks, they ranked readability as a more important criterion than the depth or the breadth of the textbook (Stang, 1975). Other studies have examined objective textbook characteristics such as number of illustrations or page length (Quershehi, 1981; Weiten, 1988). Griggs and colleagues (Griggs, 1999; Griggs et al., 1999; Griggs & Koenig, 2001) recently took a careful look at introductory psychology textbooks, including an examination of available features and levels of difficulty.

Jamison, Searle, Gilda, and Heyneman (1981) concluded that “widespread availability of textbooks in the United States preceded research on the effectiveness of instructional materials” (p. 556). This study attempted to, in part, remedy that situation. We used a list of pedagogical aids and publisher practices to determine which attributes influence textbook selection and which influence student learning (others have also studied pedagogical aids—e.g., see Marek, Griggs, & Christopher, 1999). This study expands on past research by directly asking introductory psychology instructors about textbook selection practices, particularly in the context of instructor experience in selecting textbooks.

Method

Participants

We surveyed instructors of the introductory/general psychology course. The American Psychological Association’s Office of Research provided a listing of 1,871 undergraduate and graduate psychology departments in the United States. We sent our survey to the department head of approximately every third department on the list (N = 626), and 127 departments responded (20.2% response rate).

Materials

We generated a four-page survey based on the literature, pilot testing, and consultation with content experts. The bulk of the questions were 79 different attributes presented alphabetically to introductory psychology teachers, who rated each attribute in terms of its importance in (a) selecting a textbook and (b) facilitating student learning. Subsequent demographic questions asked about the institution, the department, the introductory course, length of time teaching, and the student body. Finally, three open-ended items asked for (a) what is most needed in introductory psychology texts, (b) what needs to be eliminated, and (c) a narrative describing the decision-making procedure actually used in the textbook adoption process.

Procedure

The mailing included a cover letter; the four-page survey; a postcard; and a large, self-addressed business-reply mail envelope. We addressed the cover letter to the department chair and asked that person to forward the packet to an introductory psychology instructor. The postcard asked participants to indicate whether they would like to receive the results from the survey (a self-addressed business-reply postcard that could be mailed separate from the survey results to preserve anonymity). We mailed surveys in February 1998 with the request to reply by April 17, 1998.

Results

Demographics

Responses from the participants indicated average size of school was 6,980 students (SD = 8,078); 46.1% were from public schools and 53.9% from private schools. Similarly, 52.0% of our respondents were from a department with a graduate program in psychology, whereas 48.0% were exclusively undergraduate departments. The number of psychology majors averaged 277.2 (SD = 273.7), full-time faculty averaged 23.2 positions (SD = 34.6), part-time positions averaged 5.6 (SD = 7.1), and the number of students annually enrolled in introductory psychology courses averaged 668.0 (SD = 779.4). The typical course size of introductory psychology was reported as 75.8 (SD = 75.5). The average number of years our participants had been teaching the course was 13.9 years (SD = 9.3); on average 5.3 different instructors teach the course at their institution (SD = 7.1).

Respondents estimated that 87.1% of their students have reliable access to computers and that 84.1% have reliable access to the Internet. When asked if instructors had ever asked students for feedback on a textbook the class was using, 91.3% reported yes, and 68.3% of respondents reported that students had complained about a textbook used by the instructor. We also asked instructors if they would consider using a textbook with longer but fewer chapters (59.7% said yes) and if they would consider using a textbook with shorter but more chapters (69.9% said yes).

Descriptive Data

Table 1 provides the means for the 79 attributes on their importance in textbook selection and facilitation of student learning, respectively, ranked by the textbook selection mean. Respondents used the scale points of 0 (not at all important), 1 (slightly important), 2 (moderately important), and 3 (extremely important). Clearly, the listings are similar but not identical. For example, factors such as availability of ancillary package and computerized test item files (testbanks) were highly rated with respect to textbook selection (Ms = 2.25 and 2.24, respectively), but not on their importance in student learning (Ms = 1.72 and 0.78, respectively). On the other hand, there is appears to be good correspondence between other items in Table 1. Accuracy, readability/writing quality, and use of examples appeared in the top three in both listings.

Teaching Experience

We correlated textbook selection and facilitating student learning attributes with the question “How long have you taught introductory psychology?” (see Table 2). Length of time teaching the course correlated with 17 textbook selection attributes (15 negative correlations) and 10 facilitating student learning attributes (9 negative correlations).
We also asked participants about (a) additions to introductory psychology texts, (b) deletions from texts, and (c) the textbook adoption process at their institution. Trained coders content analyzed each question separately; raters achieved interrater reliabilities of greater than 80% for the coding scheme developed for each question.

### Table 1. Means for 79 Attributes Rated on Importance of Textbook Selection and Student Learning

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Textbook Selection M</th>
<th>Student Learning M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>2.93</td>
<td>2.68</td>
</tr>
<tr>
<td>Readability/writing quality</td>
<td>2.89</td>
<td>2.91</td>
</tr>
<tr>
<td>Examples</td>
<td>2.60</td>
<td>2.77</td>
</tr>
<tr>
<td>Currency of research</td>
<td>2.59</td>
<td>2.06</td>
</tr>
<tr>
<td>Research base</td>
<td>2.54</td>
<td>2.27</td>
</tr>
<tr>
<td>Definition of key terms</td>
<td>2.39</td>
<td>2.63</td>
</tr>
<tr>
<td>Diagrams/figures</td>
<td>2.29</td>
<td>2.33</td>
</tr>
<tr>
<td>Availability of ancillary package</td>
<td>2.25</td>
<td>1.72</td>
</tr>
<tr>
<td>Computerized test item/file/textbook</td>
<td>2.24</td>
<td>0.78</td>
</tr>
<tr>
<td>Student performance using textbook</td>
<td>2.22</td>
<td>2.06</td>
</tr>
<tr>
<td>Critical thinking exercises</td>
<td>2.20</td>
<td>2.20</td>
</tr>
<tr>
<td>Glossary</td>
<td>2.06</td>
<td>2.25</td>
</tr>
<tr>
<td>Graphs</td>
<td>2.06</td>
<td>2.08</td>
</tr>
<tr>
<td>Appealing layout of text</td>
<td>2.04</td>
<td>2.34</td>
</tr>
<tr>
<td>Student complaints about current textbook</td>
<td>2.03</td>
<td>1.74</td>
</tr>
<tr>
<td>Neuroscience and genetics section</td>
<td>2.02</td>
<td>1.75</td>
</tr>
<tr>
<td>End of chapter summaries</td>
<td>2.00</td>
<td>2.32</td>
</tr>
<tr>
<td>Customer service/publisher support</td>
<td>1.99</td>
<td>0.63</td>
</tr>
<tr>
<td>Instructor’s manual</td>
<td>1.96</td>
<td>1.02</td>
</tr>
<tr>
<td>Cost of textbook</td>
<td>1.96</td>
<td>1.40</td>
</tr>
<tr>
<td>Cultural diversity examples</td>
<td>1.95</td>
<td>1.85</td>
</tr>
<tr>
<td>Review questions</td>
<td>1.91</td>
<td>2.19</td>
</tr>
<tr>
<td>Overhead transparencies</td>
<td>1.90</td>
<td>1.86</td>
</tr>
<tr>
<td>Access to original sources of information</td>
<td>1.89</td>
<td>1.39</td>
</tr>
<tr>
<td>End-of-chapter key terms</td>
<td>1.89</td>
<td>2.24</td>
</tr>
<tr>
<td>Biological emphasis</td>
<td>1.86</td>
<td>1.50</td>
</tr>
<tr>
<td>Key terms defined in margins</td>
<td>1.85</td>
<td>2.12</td>
</tr>
<tr>
<td>Chapter opening objectives</td>
<td>1.84</td>
<td>1.97</td>
</tr>
<tr>
<td>Gender coverage</td>
<td>1.83</td>
<td>1.72</td>
</tr>
<tr>
<td>Chapter outlines</td>
<td>1.82</td>
<td>2.10</td>
</tr>
<tr>
<td>Historical information</td>
<td>1.80</td>
<td>1.59</td>
</tr>
<tr>
<td>Video series availability</td>
<td>1.80</td>
<td>1.78</td>
</tr>
<tr>
<td>Chapter length</td>
<td>1.79</td>
<td>1.98</td>
</tr>
<tr>
<td>Chapter learning objectives</td>
<td>1.78</td>
<td>1.96</td>
</tr>
<tr>
<td>Ability of chapters to stand alone</td>
<td>1.77</td>
<td>1.53</td>
</tr>
<tr>
<td>Discussion questions</td>
<td>1.76</td>
<td>1.88</td>
</tr>
<tr>
<td>Textbook length</td>
<td>1.71</td>
<td>1.51</td>
</tr>
<tr>
<td>Tables</td>
<td>1.71</td>
<td>1.78</td>
</tr>
<tr>
<td>Stories or vignettes</td>
<td>1.68</td>
<td>2.02</td>
</tr>
<tr>
<td>Color pictures</td>
<td>1.67</td>
<td>1.79</td>
</tr>
<tr>
<td>Study guides</td>
<td>1.67</td>
<td>1.98</td>
</tr>
<tr>
<td>End of section self-checks</td>
<td>1.65</td>
<td>2.10</td>
</tr>
<tr>
<td>Authors’ education and credentials</td>
<td>1.60</td>
<td>0.79</td>
</tr>
<tr>
<td>Authors’ reputation</td>
<td>1.59</td>
<td>0.65</td>
</tr>
<tr>
<td>Marginal glossary</td>
<td>1.58</td>
<td>1.73</td>
</tr>
<tr>
<td>Sidebars/marginal material</td>
<td>1.57</td>
<td>1.69</td>
</tr>
<tr>
<td>Highlighted print</td>
<td>1.54</td>
<td>1.84</td>
</tr>
<tr>
<td>Practice test questions</td>
<td>1.48</td>
<td>2.03</td>
</tr>
<tr>
<td>Inclusion of evolutionary theory</td>
<td>1.45</td>
<td>1.27</td>
</tr>
<tr>
<td>Boldface type</td>
<td>1.41</td>
<td>1.71</td>
</tr>
<tr>
<td>Lecture suggestions</td>
<td>1.40</td>
<td>0.86</td>
</tr>
<tr>
<td>CD–ROM support materials</td>
<td>1.35</td>
<td>1.37</td>
</tr>
<tr>
<td>Peer recommendations</td>
<td>1.31</td>
<td>0.51</td>
</tr>
<tr>
<td>Photo/art program</td>
<td>1.30</td>
<td>1.29</td>
</tr>
<tr>
<td>Collaboration exercises for students</td>
<td>1.25</td>
<td>1.57</td>
</tr>
<tr>
<td>Student software</td>
<td>1.21</td>
<td>1.31</td>
</tr>
<tr>
<td>Interactive CD–ROM</td>
<td>1.14</td>
<td>1.36</td>
</tr>
<tr>
<td>Annotated instructor’s edition</td>
<td>1.13</td>
<td>0.61</td>
</tr>
<tr>
<td>Suggested readings</td>
<td>1.09</td>
<td>0.98</td>
</tr>
<tr>
<td>Sales representative rapport</td>
<td>1.08</td>
<td>0.18</td>
</tr>
<tr>
<td>Charts</td>
<td>1.94</td>
<td>1.94</td>
</tr>
</tbody>
</table>

### Table 1 (Continued)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Textbook Selection M</th>
<th>Student Learning M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated Web package</td>
<td>1.08</td>
<td>1.16</td>
</tr>
<tr>
<td>Internet projects</td>
<td>0.99</td>
<td>1.15</td>
</tr>
<tr>
<td>Softcover book</td>
<td>0.95</td>
<td>0.62</td>
</tr>
<tr>
<td>Cartoons</td>
<td>0.94</td>
<td>1.19</td>
</tr>
<tr>
<td>Publisher Web site</td>
<td>0.90</td>
<td>0.69</td>
</tr>
<tr>
<td>Publisher reputation</td>
<td>0.83</td>
<td>0.25</td>
</tr>
<tr>
<td>Teaching assistant/adjunct guide</td>
<td>0.83</td>
<td>0.57</td>
</tr>
<tr>
<td>ESL study guide</td>
<td>0.83</td>
<td>1.15</td>
</tr>
<tr>
<td>Sales representative loyalty</td>
<td>0.80</td>
<td>0.11</td>
</tr>
<tr>
<td>Hardcover book</td>
<td>0.75</td>
<td>0.38</td>
</tr>
<tr>
<td>Publisher incentives</td>
<td>0.67</td>
<td>0.24</td>
</tr>
<tr>
<td>Two-column format</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>Laserdisc availability</td>
<td>0.63</td>
<td>0.65</td>
</tr>
<tr>
<td>One-column format</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>Chat room capability</td>
<td>0.45</td>
<td>0.57</td>
</tr>
<tr>
<td>Publisher loyalty</td>
<td>0.44</td>
<td>0.09</td>
</tr>
<tr>
<td>Audiotapes for student studying</td>
<td>0.37</td>
<td>0.83</td>
</tr>
<tr>
<td>Number of textbook authors</td>
<td>0.19</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Note. Scale values: 0 = not at all important; 1 = slightly important; 2 = moderately important; 3 = extremely important. Items are ranked by the textbook selection mean rating.

### Table 2. Significant Correlations* Between Length of Time Teaching Introductory Psychology and Attributes Rated on Textbook Selection and Facilitation of Student Learning

<table>
<thead>
<tr>
<th>Textbook Selection</th>
<th>r</th>
<th>Facilitating Student Learning</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture suggestions</td>
<td>−0.31</td>
<td>Lecture suggestions</td>
<td>−0.27</td>
</tr>
<tr>
<td>Instructor’s manual</td>
<td>−0.27</td>
<td>Review questions</td>
<td>−0.26</td>
</tr>
<tr>
<td>Annotated instructor’s edition</td>
<td>−0.26</td>
<td>Cultural diversity</td>
<td>−0.26</td>
</tr>
<tr>
<td>Stories or vignettes</td>
<td>−0.25</td>
<td>Critical thinking exercises</td>
<td>−0.22</td>
</tr>
<tr>
<td>Critical thinking exercises</td>
<td>−0.22</td>
<td>End of section</td>
<td>−0.50</td>
</tr>
<tr>
<td>Cultural diversity</td>
<td>−0.21</td>
<td>Self-checks</td>
<td>−0.21</td>
</tr>
<tr>
<td>Graphs</td>
<td>0.22</td>
<td>Critical thinking exercises</td>
<td>−0.18</td>
</tr>
<tr>
<td>Peer recommendations</td>
<td>−0.21</td>
<td>Currency of research</td>
<td>−0.18</td>
</tr>
<tr>
<td>Diagrams/figures</td>
<td>0.21</td>
<td>Highlighted print</td>
<td>−0.18</td>
</tr>
<tr>
<td>Collaboration exercises for students</td>
<td>−0.20</td>
<td>Graphs</td>
<td>0.18</td>
</tr>
<tr>
<td>End of section self-checks</td>
<td>−0.20</td>
<td>Highlighted print</td>
<td>−0.20</td>
</tr>
<tr>
<td>Highlighted print</td>
<td>−0.20</td>
<td>Practice test questions</td>
<td>−0.20</td>
</tr>
<tr>
<td>Practice test questions</td>
<td>−0.20</td>
<td>Currency of research</td>
<td>−0.20</td>
</tr>
<tr>
<td>Sidebars/marginal material</td>
<td>−0.20</td>
<td>Discussion questions</td>
<td>−0.17</td>
</tr>
<tr>
<td>Discussion questions</td>
<td>−0.17</td>
<td>Readability/writing quality</td>
<td>−0.17</td>
</tr>
</tbody>
</table>

Note. N = 127. *p < .05.

### Open-Ended Responses

We also asked participants about (a) additions to introductory psychology texts, (b) deletions from texts, and (c) the textbook adoption process at their institution. Trained coders content analyzed each question separately; raters achieved interrater reliabilities of greater than 80% for the coding scheme developed for each question.
The five most frequent themes about what instructors want to see added to introductory psychology textbooks were (a) greater application of the material to a personal level and relating to current issues (14.0%); (b) greater emphasis on research and experiments (12.1%); (c) Web site suggestions, computer simulations, and interactive computer programs (10.3%); (d) better integration of content (8.4%); and (e) shorter chapters (6.5%). The top five requests for deletions from the introductory text included (a) chapters on marginal topics (14.0%); (b) cartoons (10.3%); (c) some of the boxes and definitions in margins (6.5%); (d) too much depth and detail (5.6%); and (e) nonfunctional photos (4.7%). The most common process of textbook selection was individual instructor selection (64.5%); next most common was by committee vote or consensus (30.8%).

Discussion

Experience in teaching the introductory course matters when it comes to textbook selection. With regard to the attributes important to textbook selection, more experienced instructors placed greater importance on diagrams/figures and graphs than less experienced instructors. Conversely, instructors with less experience in teaching the course seemed to place higher importance on the ancillary/supplementary materials available, such as the annotated instructor’s edition, different exercises (collaboration, critical thinking, cultural diversity), discussion questions, lecture suggestions, and overhead transparencies. When examining the attributes about the facilitation of student learning, the pattern of importance ratings with course experience was remarkably similar, with more experienced instructors placing greater importance on graphs for student learning and less experienced instructors placing greater importance on the ancillary materials. It may be that as instructors teach the course more and grow in confidence, there is less reliance on the supplements and ancillaries and more attention paid to the content details of the textbook.

Optimum textbook selection serves multiple goals: Whereas the student must ultimately use the textbook, the instructor or a committee chooses a textbook. Thus, the textbook (and the entire publishing package) must appeal to the instructor before adoption and be used and valued by the student after adoption. This two-tiered process appears to be a complicated one; examination of the comprehensive list of criteria presented in Table 1 shows that although there was good correspondence between textbook selection and facilitating student learning, it was not perfect. In particular, the textbook selection process focused more on support materials available from the publisher (e.g., ancillaries, test item files), whereas the student learning process attributes focused more on the availability of student study aids, feedback from students and others, and page layout. Examination of individual items in Table 1 should provide a comprehensive picture of the relative value of pedagogical aids and publisher practices in the roles of text selection and student learning. When asked about what to add or eliminate from textbooks, instructors wanted more application of the material to a student’s life and greater emphasis on research and experiments; instructors reported most often that they would like marginal chapters eliminated and fewer cartoons.

Future research should continue to examine the textbook selection process from the point of view of the instructor decision and the impact of that decision on student learning. Certainly, the impact of emerging technologies and improved computer access may quickly change the current landscape. Because elimination of textbooks does not seem to be likely in the near future, continued study of the role that textbooks play in the postsecondary educational system should yield fruitful results. Selection of a textbook is an important decision for the introductory psychology instructor (Chatman & Goetz, 1985; Morris, 1977). That selection process is complicated by the instructor’s consideration of textbook attributes, publisher practices, and the principles that facilitate student learning, as well as the instructor’s experience in teaching the course. Fortunately, many of the practices used by faculty to select a textbook appear to be in the best interest of student learning.

References


Note

Send correspondence to R. Eric Landrum, Department of Psychology, Boise State University, 1910 University Drive, Boise, ID 83725; e-mail: elandru@boisestate.edu.
Using a Core Textbook for the Introductory Course

Richard A. Griggs
University of Florida

Sherri L. Jackson
Jacksonville University

Pam Marek
Anderson College

The introductory course often ends up only a hurried survey because current introductory psychology textbooks are too long for the normative 1-semester course. This problem is aggravated by the soaring costs of these texts. As a remedy for this situation, some teachers have recently proposed teaching the course without a standard textbook. In lieu of no text, we urge teachers to consider using an inexpensive core text. Adopting such a text would provide students with some exposure to the core introductory course content and still allow teachers to use other specialized texts and readings. We provide information on length, cost, content, organization, and pedagogical programs for the available core texts and discuss some caveats with respect to using such texts.

A continuing criticism of standard introductory psychology textbooks is that they are too long (e.g., Landrum, 2000; Nallan, 1997). The mean length of current texts is 774 pages (Griggs, Jackson, Christopher, & Marek, 1999). Even the briefest versions of these texts are long, with a mean length of 644 pages (Griggs & Koenig, 2001). Because 89% of introductory psychology courses are only one term in length (Miller & Gentile, 1998), it would appear that teachers could not completely cover these lengthy texts. Indeed, Miller and Gentile found this to be the case in their national survey of introductory psychology teachers. The teachers reported covering an average of 68% of the text topics. In addition, they reported that the course most often ends up being a rather hurried survey and that they are not meeting other important goals such as engaging students in scientific inquiry about psychological processes and developing students' critical thinking skills. This situation is made even more problematic by the cost of introductory textbooks. Like all textbooks, introductory texts have soared in price (Freberg, 1999), with texts selling in the $80 to $90 range and briefer versions in the $50 to $60 range. Given these prices and the inability of teachers to assign the entire text, some students may not even purchase texts for the introductory course, as Sikorski et al. (2002) found.

As a remedy for this situation, some teachers have recently proposed teaching the introductory course without a standard text (Wulff, 1999). Such proposals usually involve using more specialized books and readings. In some cases, the teacher makes a set of standard texts available at some central location (e.g., on library reserve) for students to use as references for concept definition and elaboration (e.g., Zechmeister & Zechmeister, 1999). In lieu of no text, we propose an alternative solution that still allows teacher flexibility in achieving teachers' individual course goals but eliminates the need for making standard textbooks available for student use and does not substantially increase text cost.

We propose adopting a core introductory psychology text. We do not mean one of the recent brief, black-and-white versions of an existing standard text that is tied to the Web for supplementary material (e.g., Lefton, 2001). Such texts are still 450 to 500 pages in length and in the $40 to $45 cost range. Core texts are both much briefer and cheaper. These texts provide comprehensive, student friendly but very concise coverage of the major topics and concepts in standard introductory textbooks. They use illustrations sparingly, mainly figures and tables only where necessary. There is no "chartjunk" (Tuft, 1983), graphic embellishment (e.g., irrelevant colored photographs) that may detract from the text. In brief, these texts present only the essential core information for the introductory course.

Introductory teachers could adopt a core text in conjunction with one or more supplementary, specialized texts or sets of readings to meet their individual course goals. For example, a teacher might want to emphasize critical thinking and thus could also adopt one or more brief critical thinking supplementary texts (for a description of many of these texts, see Marek, Jackson, Griggs, & Christopher, 1998) along with the core text. Adopting a core text as a reference resource for students would also both preclude the need for a set of standard introductory texts at some central location and facilitate student access to the core information because students would have their own books, highlighting this information. To inform interested introductory teachers, we provide information about the length, cost, content, organization, and pedagogical programs for the core review texts presently available.

Description of Available Core Review Texts

We checked the Web sites of all major publishers of psychology textbooks and online booksellers to identify the available brief core texts. We did not include psychology test preparation guides or foreign texts. We identified eight core texts and obtained copies from the publishers.

Each core text in psychology is part of a series of review books in several disciplines. These series seem to fall into three categories—review books, extended review books, and outline books. These categorical labels are somewhat arbitrary because all these texts provide core reviews and almost all use enumerated outline text format. It is mainly the extent of the review and the nature of the pedagogical program that differentiate the various types of core texts. The extended review and outline texts provide longer reviews, and they tend to use more pedagogical aids, especially the outline texts. Our examination of these texts makes these differences clearer.

There are three review, two extended review, and three outline core texts presently available. Reference, series, and price information for these eight core texts appears in the Appendix. To facilitate our discussion of these texts, we use the following publisher names as short titles to refer to them. The review texts are Cliff's, Barron's, and Research and Education Association (REA); the extended reviews are Barron's+ and REA+; and the outline texts are Harcourt, Schaum's, and HarperCollins. For most comparisons, we distinguish only between the three categories, but sometimes we point out differences between specific texts. In these cases, we use short titles.

314 Teaching of Psychology
First, we considered length and cost. The review texts are briefer than the extended review and outline texts (M = 209 pp. vs. M = 373 pp. and 296 pp., respectively). All three types are substantially shorter than the briefer versions of standard introductory texts. Similarly, the cost of all of the review texts is clearly less than that of standard introductory textbooks or their briefer versions: M = $9.93 for review texts, $16.43 for extended review texts, and $15.90 for outline texts. The mean for the review books is distorted by REA because it is published in two volumes, which doubles its cost. The mean for the other two review texts is $7.95. The costs of the extended and outline review texts are fairly comparable, about double that for the two singular review texts. However, the prices for all of these core texts are far below the prices for standard textbooks.

Next we considered topic coverage and organization. The core texts had essentially the same chapter topics and used the typical chapter organization of the standard texts. Only Cliffs, Harcourt, and HarperCollins, however, had a health psychology chapter, and only Cliffs and Harcourt had an atypical chapter (diversity and evolution/genetics, respectively). In addition, the chapter organization of Cliffs is somewhat atypical in that the "Memory," "Thought/Language," and "Intelligence" chapters are near the end of the text rather than in the middle.

To check that these texts actually do provide the core content of the introductory course, we compared the core terminology used in the three review books with that in the majority of standard introductory texts. Given their extreme brevity, we thought that these review texts might not sufficiently cover the language of introductory psychology and thus comprised the best test to check for core terminology. Using a page-by-page text analysis, we first created a computer file of all of the boldface terms in each of the three review texts and then sorted the subset that appeared in the majority of these texts (i.e., any two or all three). We next searched this subset list for each of the 421 terms found in the majority of current standard introductory texts (given in Proctor, Griggs, & Bujak, 2001). We found the vast majority, 91% (383 of 421). This finding clearly indicates that these review texts do achieve their main goal, providing concise coverage of the core concepts of introductory psychology. This result should not be too surprising, however, because some of the core text authors (e.g., Sternberg) have authored standard introductory textbooks.

Last, we examined the pedagogical programs for the core texts. We checked for all 15 aids (e.g., chapter summaries, concept checks, practice tests) examined by Marek, Griggs, and Christopher (1999) in their study of standard introductory texts (for a listing of all 15 aids, see Table 1 in Marek et al., 1999). We expected that core texts would use relatively few pedagogical aids given their emphasis on concise presentation, and this was the case: M = 1.7 for review texts, 3.0 for extended review texts, and 4.3 for outline texts. These figures are all less than the mean for standard introductory textbooks, M = 6.9 (Marek et al., 1999). Only two aids appeared in a majority (five or more) of the eight books. All but Barron's, which used italics, used boldface type for emphasis; all but Barron's, REA, and REA+ used chapter outlines/overviews. With respect to self-tests (an aid valued by students but not present in many standard texts; Marek et al., 1999), only REA+, Harcourt, and Schaum's provided such tests. Schaum's actually contained three noncumulative exams and a cumulative final exam.

In summary, although we think that adopting a core text is an effective, low-cost way for introductory teachers who do not use standard textbooks to give their students exposure to the basic course content, there are some caveats. First, we want to make it clear that we are not endorsing the use of these texts in place of standard textbooks for the introductory course. Our recommendation for adoption pertains only to cases in which more specialized books and readings are being used in lieu of a standard textbook. We do not view these core texts as adequate substitutes for standard textbooks. Second, obtaining complimentary examination copies of these core texts to consider them for course adoption will likely not be as easy as it is for standard texts. Half of the publishers of these texts (i.e., Cliffs Notes, Barron's, and REA) are not standard textbook publishers, and all of these books are classified as trade books. Thus, teachers might need to purchase examination copies. However, given their inexpensive prices, this consideration should not pose a major problem in the examination process. Third, if ordered as a textbook, your bookstore will likely mark up the price of these core texts just as they do with the prices of standard textbooks. Thus, the exact to-the-student price will be slightly more than the price we report—how much depends on the specific bookstore.

References


Vol. 29, No. 4, 2002

315
Using the Barnum Effect to Teach Psychological Research Methods

Thomas E. Boyce
University of Nevada, Reno

E. Scott Geller
Virginia Polytechnic Institute and State University

We used the Barnum Effect (Dickson & Kelly, 1985) to teach principles of research and to promote participation in 2 graduate psychological research methods courses. During the first week of class, 2 instructors presented themselves as experts in graphology and collected handwriting samples from students. Subsequently, each student received the same ambiguous one-size-fits-all personality assessment, presumably based on his or her handwriting. The students completed questionnaires about their perceptions of graphology prior to and immediately after receiving their personality assessments and 8 weeks later. Students significantly increased their ratings of graphology as a science after receiving their personality assessments, then lowered their ratings significantly after debriefing. We used the demonstration to teach various aspects of displaying and interpreting data and to highlight the pitfalls of pseudoscience.

The Barnum Effect refers to people's tendencies to accept general and ambiguous statements as descriptive of their unique personalities (Dickson & Kelly, 1985). It was named after P. T. Barnum, the famous circus entrepreneur, who claimed he owed his success to two basic principles: "There's a sucker born every minute" and "have something for everybody" (Myers, 1993, p. 153). Both of these principles seem to be verified by people's acceptance of one-size-fits-all personality statements.

Dickson and Kelly (1985), for example, administered a personality test to participants, provided them the same bogus personality profile presumably generated from the personality test they took, and then had participants rate the perceived accuracy of the profile they received. Most participants rated the general personality profile as a good or excellent description of themselves. Furthermore, participants' subjective ratings increased directly with the favorableness of the general profile.

The literature is replete with examples of applications of the Barnum Effect to teach ethics in research (Beins, 1993) and to identify variables that moderate its effect (for a review, see Dickson & Kelly, 1985). However, we found no published research on the use of the Barnum Effect to promote healthy skepticism of pseudoscience and teach university students research design and methodology. Here we document the results of a Barnum procedure designed to accomplish these objectives.

We selected graphology as the personality assessment procedure because of people's tendencies to assign higher accuracy ratings to profiles generated by more ambiguous assessment procedures (Dickson & Kelly, 1985). Our aims were to demonstrate the allure of and reduce students' belief in pseudoscience as well as develop an appreciation for em-
Recent research on introductory psychology textbooks has indicated that these texts do not share a common core vocabulary. For example, Zechmeister and Zechmeister’s (2000) content analysis of the glossaries of 10 introductory textbooks revealed that only 64 of 2,505 different glossary terms appeared in all 10 glossaries and that about half of the concepts appeared in only 1 glossary. Similarly, Landrum (1993) conducted a page-by-page content analysis of “important” terms in six introductory textbooks and found that 1,600 of 2,742 different terms were unique and only 126 were in all six textbooks. Analyzing 52 introductory textbook indexes, Quereshi (1993) also found little commonality; only 3 terms appeared in all of the indexes and only 141 appeared in 75% of the indexes. Thus, three studies, each employing a different type of vocabulary analysis, found that introductory texts did not share a very large common vocabulary. Zechmeister and Zechmeister (2000) pointed out that this lack of a common core vocabulary in introductory textbooks poses a real problem for introductory teachers trying to teach the course via the prescription “less is more,” because there is no clear definition of what the “less” should be.

Zechmeister and Zechmeister (2000) further pointed out that the lack of agreement across studies exacerbates this problem. For example, of the 126 core terms that Landrum found in his 6-book sample, only 44% were in even 8 or more of the 10 textbooks in the Zechmeister and Zechmeister sample. As Zechmeister and Zechmeister concluded, “this lack of convergent validity is obvious and troublesome” (p. 9). However, the sampled textbooks varied greatly across the different studies. To obtain more definitive data on the core vocabulary for introductory textbooks, we analyzed the glossaries of the current population of introductory textbooks. Thus, a teacher who wanted to use the common core vocabulary in introductory texts as the definition of “less” for teaching the course could do so. We used text glossaries because, as Zechmeister and Zechmeister argued, these terms are the ones text authors have identified as most important and students as the most likely to be tested. Landrum (1993) enumerated several other possible uses for these common core data. Of most relevance to this study, teachers could use the common core vocabulary as an aid in the textbook evaluation and selection process. Given the importance of this function, we computed not only how well each text covers the common core vocabulary but also the uniqueness of each text’s vocabulary.

Landrum and Hormel (2002) found that introductory psychology teachers rated highly both a text’s definition of terms and its glossary when they selected a text and established goals for student learning. For example, teachers rated these text attributes fourth and seventh, respectively, out of 79 attributes for importance to student learning. Chatman and Goetz (1985) also suggested using the extent to which textbooks cover key concepts as a means of reducing the size of the set of texts chosen for further, more detailed analyses during textbook evaluation. For introductory teachers who might heavily weight common core coverage and uniqueness of vocabulary in their text evaluation process, our data should prove extremely beneficial.

Method

In their compendium of introductory psychology textbooks, Koenig, Griggs, Marek, and Christopher (2003) listed 35 current (latest copyright from 2000 to 2003) introductory textbooks. The only other current introductory textbooks are briefer versions of some of these 35. One of the 35 current textbooks, Gaulin and McBurney (2001), was not appropriate for our study. It is idiosyncratically structured around evolution and its applications to psychology and employs a very atypical vocabulary and thus is not comparable to normal introductory texts. Because introductory textbooks are on a 3-year revision cycle, Koenig et al.’s (2003) compendium included only texts with a latest copyright date within the last 4 years (i.e., those textbooks reasonably viable for adoption consideration). To identify additional textbooks that have not been revised recently but might possibly be in the future, we checked the Web sites of introductory textbook publishers for texts with the latest copyright as old as 1997. This search yielded 10 additional textbooks with the
latest copyright from 1997 to 2003 (a time period greater than twice the typical text revision cycle). Reference information for these 44 textbooks appears in the Appendix.

After entering each textbook’s glossary of terms into a computer database, we identified synonyms (e.g., *double-blind procedure* and *double-blind study; conditioned reinforcer* and *secondary reinforcer*) through discussion and by consulting Corsini’s (1999) psychology dictionary. We counted the frequency of each term across the 44 texts and also how many different terms occurred in all 44 texts, in 43 texts, and so forth, down to unique terms.

Based on Zechmeister and Zechmeister (2000), we employed a criterion of inclusion in 80% (i.e., 35) or more of the 44 text glossaries as the initial common vocabulary criterion. Once we determined this set of terms, we compared it to the list of terms included in 8 or more of the 10 textbook glossaries in Zechmeister and Zechmeister’s textbook sample (obtained from J. S. Zechmeister, personal communication, October 2000). To provide teachers with a larger set of common core terms to use in their teaching, we also used a more lax commonality criterion: terms appearing in more than 22 of the 44 text glossaries. To aid teachers in text evaluation and selection, we computed the percentage of terms in each of the 44 texts using each criterion. In addition, we computed the uniqueness of each text’s glossary by determining the number of terms appearing only in that text’s glossary. Because uniqueness is positively correlated with glossary size (Zechmeister & Zechmeister, 2000), we also counted the total number of terms appearing in each text’s glossary.

Results and Discussion

After combining similar terms within and between text glossaries, the total number of different glossary terms across the 44 textbooks was 6,269. As in previous studies, the commonality between text glossaries was extremely low. Only 14 (0.22%) of the 6,269 terms appeared in all 44 glossaries. Over half (3,446; 55%) were unique to only 1 text glossary, and about 74% (4,654) were in 3 or fewer text glossaries.

Only 155 terms (2.5%) met the commonality criterion of appearing in 80% or more of the text glossaries. These terms appear in Table 1, grouped by their frequency of glossary inclusion from 44 to 35. More of these terms come from the biological psychology (25) and learning (25) chapters than any of the other standard introductory chapters. These two chapter topics accounted for almost one third (32%) of the common core terms. The number of terms for any other chapter topic was less than 10%, except for the introductory—methods chapter, which accounted for 10%.

The size of the common core set of terms (155) is less than the 197 found by Zechmeister and Zechmeister (2000) for their sample of 10 texts, but this finding is not surprising because we analyzed the 44-text population and not a small sample. Of the 197 terms included in 80% of Zechmeister and Zechmeister’s text sample, 64 (32%) were not in our common core. In addition, 22 other terms were part of our common core but not theirs. The less restrictive criterion of inclusion in more than 50% of the texts yielded 415 terms (6.6% of the 6,269 terms), which means that over 93% of the
total different glossary terms do not appear in even 50% of the texts.

Table 2 presents the number of glossary terms, the percentage of the 155 (in 80% or more of the texts) and 415 (in more than 50% of the texts) common core terms included in the glossary, and the glossary’s uniqueness percentage for each of the 44 texts. With respect to glossary size, the glossaries ranged from only 303 terms to 1,551 terms. The median glossary size was 682 terms; only 9 texts included more than 800 terms and 6 texts had fewer than 500 terms. Coverage of the set of 155 common core terms ranged from 47.7% to 98.1%, with median coverage of 91.6%. Thus, the typical text included most of the core terms, which would be expected given the core inclusion criterion of appearing in 80% or more of the texts.

There was more variability among texts in their coverage of the set of terms derived from the criterion of appearing in more than 50% of the glossaries. Such coverage ranged from 26.5% to 98.1%, with median coverage of 73.2%. Glossary uniqueness ranged from 1.4% to 29.7%, with a median of 10.6%. As expected, the one-tailed Pearson product-moment correlation between these uniqueness percentages and the texts’ glossary sizes was significant, $r(44) = .54, p < .01$. Thus, teachers considering texts with larger glossaries should be expected given the core inclusion criterion of appearing in 80% or more of the texts.

There was more variability among texts in their coverage of the set of terms derived from the criterion of appearing in more than 50% of the glossaries. Such coverage ranged from 26.5% to 98.1%, with median coverage of 73.2%. Glossary uniqueness ranged from 1.4% to 29.7%, with a median of 10.6%. As expected, the one-tailed Pearson product-moment correlation between these uniqueness percentages and the texts’ glossary sizes was significant, $r(44) = .54, p < .01$. Thus, teachers considering texts with larger glossaries should be expected given the core inclusion criterion of appearing in 80% or more of the texts.

Table 1 (Continued)
7.2% uniqueness) or a relatively small glossary with a large uniqueness percentage (e.g., the Schlinger & Poling text with 303 terms but 21.6% uniqueness).

In summary, we have identified the common core vocabulary among current introductory textbooks. As suggested by recent studies examining samples of introductory textbooks, the common core is not very large. However, teachers may use these common core data in structuring their individual courses to ensure coverage of core vocabulary. A teacher who wants to cover a more substantial common core can use the set of 415 terms derived from the simple majority criterion. Individual introductory textbooks vary greatly not only in their coverage of the common core but also in their glossary size and uniqueness. Teachers may want to include these glossary size, coverage, and uniqueness data in their introductory textbook evaluation and selection process to reduce the number of texts that they evaluate more thoroughly.

References

### Table 2. Number of Glossary Terms, Percentage Inclusion of Common Core Terms (Both in More Than 50% and 80% or More of the Texts), and Percentage Glossary Uniqueness for each Introductory Textbook

<table>
<thead>
<tr>
<th>Textbooka</th>
<th>No. of Glossary Terms</th>
<th>% Terms in More Than 50% of the Textsb</th>
<th>% Terms in 80% or More of the Textsc</th>
<th>% Glossary Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barker</td>
<td>524</td>
<td>62.7</td>
<td>85.2</td>
<td>12.5</td>
</tr>
<tr>
<td>Baron</td>
<td>663</td>
<td>71.8</td>
<td>91.6</td>
<td>6.9</td>
</tr>
<tr>
<td>Bernstein et. al.</td>
<td>648</td>
<td>76.1</td>
<td>91.6</td>
<td>6.9</td>
</tr>
<tr>
<td>Bourne &amp; Russo</td>
<td>759</td>
<td>76.1</td>
<td>92.3</td>
<td>10.9</td>
</tr>
<tr>
<td>Carlson &amp; Buskist</td>
<td>787</td>
<td>75.9</td>
<td>92.3</td>
<td>14.6</td>
</tr>
<tr>
<td>Coon</td>
<td>1551</td>
<td>83.9</td>
<td>98.1</td>
<td>29.7</td>
</tr>
<tr>
<td>Davis &amp; Palladino</td>
<td>648</td>
<td>77.1</td>
<td>93.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Feldman</td>
<td>496</td>
<td>65.1</td>
<td>90.3</td>
<td>6.0</td>
</tr>
<tr>
<td>Fernald</td>
<td>911</td>
<td>83.9</td>
<td>96.1</td>
<td>14.4</td>
</tr>
<tr>
<td>Franzoi</td>
<td>629</td>
<td>81.7</td>
<td>96.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Gazzaniga &amp; Heatherton</td>
<td>579</td>
<td>62.7</td>
<td>82.6</td>
<td>10.9</td>
</tr>
<tr>
<td>Gerow &amp; Bordens</td>
<td>494</td>
<td>63.6</td>
<td>86.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Gerrig &amp; Zimbardo</td>
<td>645</td>
<td>72.0</td>
<td>89.7</td>
<td>10.8</td>
</tr>
<tr>
<td>Gleitman et al.</td>
<td>1186</td>
<td>77.6</td>
<td>84.5</td>
<td>23.2</td>
</tr>
<tr>
<td>Gray</td>
<td>554</td>
<td>58.1</td>
<td>80.0</td>
<td>13.9</td>
</tr>
<tr>
<td>Hockenbury &amp; Hockenbury</td>
<td>709</td>
<td>81.9</td>
<td>98.1</td>
<td>7.3</td>
</tr>
<tr>
<td>Hoffman</td>
<td>479</td>
<td>73.3</td>
<td>92.9</td>
<td>2.9</td>
</tr>
<tr>
<td>Kalat</td>
<td>703</td>
<td>72.3</td>
<td>87.1</td>
<td>10.8</td>
</tr>
<tr>
<td>Kassin</td>
<td>536</td>
<td>74.9</td>
<td>88.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Kosslyn &amp; Rosenberg</td>
<td>743</td>
<td>78.1</td>
<td>94.2</td>
<td>13.1</td>
</tr>
<tr>
<td>Lahey</td>
<td>671</td>
<td>70.4</td>
<td>91.0</td>
<td>11.5</td>
</tr>
<tr>
<td>Lefton &amp; Brannon</td>
<td>528</td>
<td>69.6</td>
<td>89.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Matlin</td>
<td>784</td>
<td>78.8</td>
<td>94.2</td>
<td>12.0</td>
</tr>
<tr>
<td>Morris &amp; Maisto</td>
<td>688</td>
<td>76.9</td>
<td>94.2</td>
<td>4.9</td>
</tr>
<tr>
<td>Myers</td>
<td>553</td>
<td>78.6</td>
<td>94.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Mynatt &amp; Doherty</td>
<td>312</td>
<td>31.6</td>
<td>52.3</td>
<td>18.7</td>
</tr>
<tr>
<td>Nairne</td>
<td>529</td>
<td>78.6</td>
<td>91.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Nevid</td>
<td>915</td>
<td>87.5</td>
<td>92.3</td>
<td>7.2</td>
</tr>
<tr>
<td>Passer &amp; Smith</td>
<td>756</td>
<td>77.8</td>
<td>96.1</td>
<td>11.2</td>
</tr>
<tr>
<td>Pettijohn</td>
<td>1041</td>
<td>88.0</td>
<td>96.1</td>
<td>12.1</td>
</tr>
<tr>
<td>Plotnik</td>
<td>991</td>
<td>82.2</td>
<td>94.8</td>
<td>17.0</td>
</tr>
<tr>
<td>Rathus</td>
<td>825</td>
<td>81.7</td>
<td>96.8</td>
<td>10.3</td>
</tr>
<tr>
<td>Santrock</td>
<td>722</td>
<td>68.2</td>
<td>89.0</td>
<td>3.6</td>
</tr>
<tr>
<td>Schlinger &amp; Poling</td>
<td>303</td>
<td>26.5</td>
<td>47.7</td>
<td>21.6</td>
</tr>
<tr>
<td>Sdorow &amp; Rickabaugh</td>
<td>704</td>
<td>83.9</td>
<td>95.5</td>
<td>7.0</td>
</tr>
<tr>
<td>Smith</td>
<td>758</td>
<td>73.0</td>
<td>88.4</td>
<td>11.0</td>
</tr>
<tr>
<td>Smith et al.</td>
<td>773</td>
<td>65.5</td>
<td>89.0</td>
<td>21.6</td>
</tr>
<tr>
<td>Sternberg</td>
<td>971</td>
<td>83.4</td>
<td>94.8</td>
<td>17.1</td>
</tr>
<tr>
<td>Uba &amp; Huang</td>
<td>550</td>
<td>71.8</td>
<td>90.3</td>
<td>5.2</td>
</tr>
<tr>
<td>Wade &amp; Tavis</td>
<td>428</td>
<td>60.7</td>
<td>83.2</td>
<td>5.6</td>
</tr>
<tr>
<td>Weiten</td>
<td>693</td>
<td>73.3</td>
<td>91.6</td>
<td>6.3</td>
</tr>
<tr>
<td>Westen</td>
<td>916</td>
<td>83.9</td>
<td>96.1</td>
<td>14.4</td>
</tr>
<tr>
<td>Wood &amp; Wood</td>
<td>773</td>
<td>8.12</td>
<td>93.5</td>
<td>5.4</td>
</tr>
<tr>
<td>Wortman et al.</td>
<td>660</td>
<td>73.0</td>
<td>87.7</td>
<td>9.7</td>
</tr>
</tbody>
</table>

#The Appendix provides reference information for all 44 texts. #Percentage of the 415 terms in the glossaries of more than 50% of the 44 textbooks included in each text. #Percentage of the 155 terms in the glossaries of 80% or more of the 44 textbooks included in each text. #Percentage of a textbook’s glossary terms not appearing in any other text glossary.


Appendix

44 Introductory Textbooks Used in Study


Notes

1. The listings of the glossary terms included in more than 50% of the 44 text glossaries or any other reported analysis are available from the authors on request.

2. We thank Jeanne Zechmeister for providing us with the core concept data from Zechmeister and Zechmeister (2000), and Randolph Smith and three anonymous reviewers for their valuable comments on an earlier version of this article.

3. Send correspondence to Richard A. Griggs, Department of Psychology, P.O. Box 112250, University of Florida, Gainesville, FL 32611; e-mail: rgriggs@ufl.edu.

Conscientiousness Is Key: Incentives for Attendance Make Little Difference

Maureen A. Conard
Sacred Heart University

This study examined differences in class attendance at different levels of conscientiousness and incentives (3.5% vs. 6% of course points). Results of a 2 × 2 (Level of Incentives × Level of Conscientiousness) ANOVA indicated a significant main effect for conscientiousness. Conscientiousness accounted for 14% of the variance in attendance, compared to 1% for incen-
TOPICAL ARTICLES

Evaluating the Electronic Textbook: Is It Time to Dispense With the Paper Text?

James A. Shepperd  
University of Florida

Jodi L. Grace  
St. Thomas University

Erika J. Koch  
St. Francis Xavier University

We examined the perceptions and performance of students who used an electronic versus a traditional paper textbook. Introductory psychology students (N = 392) who chose between the 2 formats did not differ in course grades. However, students using the electronic text reported spending less time reading for class compared to students using the paper text and generally evaluated the electronic text unfavorably. No student who purchased an electronic text in a prior class chose to purchase it for introductory psychology. These findings suggest that it may be premature to abandon the paper text in favor of the electronic text.

A recent innovation in computerized instruction is the electronic textbook. The electronic text is sold on compact disc or DVD and offers text, photos, graphs, and tables in digital format. Electronic textbooks offer several advantages over paper textbooks. Provided students do not print a hard copy of the text, they are less expensive, lighter, less bulky, and environmentally friendly. In addition, because the electronic text is stored on the student’s computer, it offers many of the advantages of computers such as the ability to conduct searches of terms, the ability to cut sections of the text and paste them into other documents, and may include video presentations of selected topics, and interactive maps and figures that can help students learn complex material (Blumenstyk, 2001a; Young, 2001). Students sometimes prefer the electronic medium. For example, students are more likely to consult an electronic dictionary than a paper dictionary (Aust, Kelley, & Roby, 1993).

These advantages notwithstanding, electronic textbooks have disadvantages. Perhaps most important, the electronic text requires that students own or have access to a computer. Unless students have laptops, electronic texts also can be inconvenient for students accustomed to bringing their textbook to class or reading from it during breaks between classes. Even at home and in the dorm room, students are largely restricted to studying at a desk or table. Additionally, although research reveals no difference in comprehension of material presented on paper versus electronically (Aust et al., 1993), some students and instructors complain that reading from a computer text feels disjointed (Young, 2001). Other readers report that they tend to skim electronic text material, but if they plan to read the material in its entirety, they prefer a print version (Rho & Gedeon, 2000). Also, the cut-and-paste function of electronic textbooks makes plagiarism easier, allowing students to claim the text material as their own in their papers (Blumenstyk, 2001b). Finally, students may lose access to their textbook if the computer fails.
Instructors choosing between the electronic and paper text for their classes face a more pragmatic concern: How do these different formats affect reading and understanding of the material and, ultimately, performance in the class? From a pedagogical standpoint, the ability to manipulate digital material (e.g., cutting and pasting important passages) might lead students to be more active in their reading (Blumenstyk, 2001a). In addition, students may be more engaged and attentive when reading because using the electronic textbook requires sitting at the computer.

We conducted this study to evaluate the electronic textbook. We examined whether students who elected to use an electronic versus a paper textbook in a college course differed in course grades and in amount of time spent reading. We also asked students to evaluate the electronic text.

Method

Participants

Of the 466 students enrolled in an introductory psychology course in the first week of class, 392 (98 men, 257 women, 37 sex unreported) took the final exam and completed the exit survey. Sixty percent of students completing the exit survey were first-year students, 24% sophomores, 10% juniors, and 4% seniors (2% did not report their year in school).

Materials and Procedures

The text for the class was Myers’s Psychology (2001). During the first class of the semester, a sales representative distributed a compact disc (CD) version of the text to all students. The representative provided a brief demonstration of how the electronic text worked and described how students could purchase it. The syllabus noted the costs of both versions of the textbook. The cost of the paper text was $81.25 new and $60.95 used plus an additional $22.75 if students chose to purchase the study guide that accompanies the text. The electronic text, including the study guide, cost $40. The instructor emphasized that the version of the textbook students chose to use was entirely their decision. Students who chose to purchase the text received a password allowing them to access the text on their CD.

The items on the exit survey assessed the electronic text in three categories. The first category consisted of demographic questions. To assess whether students who chose the electronic text differed from students who chose the paper text, students reported whether they had purchased the electronic text for the class, whether they had ever purchased an electronic text for another class, whether they owned a computer, and their current grade point average (GPA). In addition, students reported how knowledgeable they were about computers (1 = not at all; 7 = extremely), how much they enjoyed using computers (1 = not at all; 7 = very much), and how helpful they found the text for the class in preparing for exams (1 = not at all; 7 = extremely). Finally, students indicated where they typically studied: home or dorm, library, bookstore or coffee shop, and other. We combined the latter three choices to form a single classification that we labeled “away from home/dorm.”

The second category consisted of performance-related questions. We asked students to report how many hours per week they spent reading for the course. In addition, we obtained the grade participants actually received in the course.

The third category asked students who elected to purchase the electronic text to evaluate it. Students reported (1 = not at all and 7 = extremely) how much they liked the electronic text, its convenience, its ease of use, and their willingness to use an electronic text in future classes. In addition, students indicated whether they would recommend the electronic text to other students (1 = not at all; 7 = very much). Finally, students reported which text (paper vs. electronic) they would purchase for the introductory psychology class if they had the choice again.

Results

Background Factors

Most of the students purchased the paper text (n = 330, 90%) versus the electronic (n = 37, 10%). Statistical analysis revealed that on the whole, the two groups of students did not differ in current GPA, computer literacy, enjoyment in using computers, and how helpful students found the text in preparing for exams (all ps > .20, all η² < .01). Likewise, students who purchased the electronic versus paper text did not differ in sex, ownership of a computer, use of the optional study guide for the class, and where they studied, all χ² < 1.60, p > .05, Φ < .08.

Although students who purchased the paper versus electronic text were generally similar, two differences emerged. First, students who purchased the electronic text had, on average, more years of college education
(M = 2.10, SD = 1.12) than did students who purchased the paper text (M = 1.50, SD = .76), t(364) = 4.16, p < .001, η² = .05. Second, of the 36 students who purchased an electronic text for a prior course, not one purchased the electronic text for introductory psychology. Conversely, the 37 students who purchased the electronic textbook for introductory psychology had no prior experience using an electronic textbook, χ²(1, N = 367) = 4.48, p < .05, Φ = .11.

**Outcome Factors**

Our primary interest was whether participants who purchased the electronic versus the paper textbook differed on the various outcome measures. To address this question, we examined two measures: (a) how much time students reported reading for the class, and (b) what grade students achieved in the class.

The two groups of students differed in how much they reported studying for the class. Students who purchased the paper text reported that they studied an average of 2.3 hr per week (SD = .80), whereas students purchasing the electronic textbook reported they studied 2.0 hr per week (SD = .80), t(364) = 2.07, p < .05, η² = .01. It is possible that the electronic text somehow facilitated study and thus reduced the number of hours students needed to study. For example, the copying and pasting feature may have made the task of taking notes from the text quick and easy. Importantly, the difference in texts accounted for only 1% of the variance in time spent studying. Moreover, the number of hours that students studied per week was not significantly correlated with their final grade regardless of whether students used the paper text, r (309) = −.04, or the electronic text, r (32) = −.28, both ps > .05.

We also examined whether users of the electronic text performed better in class than did users of the paper text. We found no significant difference in final course grade between students who adopted the electronic text (M GPA = 2.3, SD = .87) versus the paper text (M GPA = 2.5, SD = 1.04), t(340) < 1.0, p > .20, η² = .00.

In summary, users of the electronic text differed from users of the paper text in how much they reported studying for the class each week. However, they did not differ significantly in the final grade they received in the class.

**Evaluations by the Users of the Electronic Text**

On a 7-point scale with higher numbers indicating more favorable ratings, students were inclined to view the electronic text as somewhat easy to use (M = 4.9, SD = 1.78) but were generally neutral in their liking for the electronic text (M = 4.0, SD = 1.96). In addition, they were unfavorable in their ratings of the convenience of the electronic text (M = 3.2, SD = 1.82), as well as in their willingness to use the electronic text in the future (M = 3.3, SD = 2.13) and to recommend the text to a friend (M = 3.3, SD = 2.05).

**Discussion**

Our findings suggest that instructors should exercise caution before adopting electronic textbooks. When we asked students who purchased the electronic textbook to evaluate it, we found that students on average were neutral in their ratings of liking. In addition, when we compared students who purchased the electronic textbook with students who purchased the paper textbook, we found that students did not report spending more time reading from the electronic textbook or receive a higher grade in the course. If anything, our data suggest reasons for concern in adopting the electronic textbook. Students who purchased the electronic textbook did not evaluate it favorably. Although they reported that the electronic text was somewhat easy to use, they appeared not to like it much, rated it as somewhat inconvenient, and seemed reluctant to recommend it to others. Only one third of the students purchasing the electronic text said that they would do so again if given a chance. Finally, all of the students who purchased an electronic text in a prior class chose not to purchase the electronic text for introductory psychology. This finding is particularly surprising given that the electronic text was notably less expensive.

Yet not all results for the electronic textbook are dire. Although students who used the electronic text reported spending less time reading for the class than did students using the paper text, the two groups did not differ significantly in the grade they received for the course. Perhaps the electronic text allows students to achieve the same grade as students using the paper text, but in less time.

We caution against drawing causal inferences from our findings. The students in this study were not randomly assigned to purchase a specific type of textbook and, thus, our findings for time spent studying are open to multiple interpretations. The electronic text may have produced greater efficiency in studying, with students absorbing more material in less time. Or, the inconvenience of the electronic text may...
have led students to use it less. Perhaps students who wanted to spend less time and effort on the class also opted to purchase the less expensive electronic text. Complicating matters further is the fact that reports of the amount of time spent studying were unrelated to the grade received in the course among both users of the paper text and users of the electronic text. Finally, our results are based on a single group of students using a single introductory text. Although we anticipate that our findings would likely replicate for other introductory psychology textbooks, it is unknown whether the findings would replicate for textbooks used in more advanced courses.

Our findings suggest that instructors should not rush to adopt the electronic text, even though publishers may offer price discounts when instructors exclusively adopt the electronic text. However, we are reluctant to sound the death knell for the electronic textbook just yet. The electronic textbook offers some impressive features such as search functions, interactive tutorials, and colorful graphics. In addition, computer programs and applications are updated and improved at a dizzying speed with new versions appearing regularly. Although users may dislike or find inadequate a particular application today, the bugs and annoying features can be corrected in the next version to appear on the market. Finally, the electronic textbook costs less than the paper text. Because the production cost is low, costing only a few cents for a CD, the price will likely fall further should the electronic text achieve wide adoption.

Perhaps some of the misgivings students expressed toward the electronic text arose from the novelty of the medium. For most people, reading a book on a computer is an experience that may require an adjustment period. We note a potential parallel in writing and using computers. Learning to use a computer to write required a learning curve. Perhaps a similar process lies ahead with electronic texts. Electronic texts may increase in popularity once students become more familiar and comfortable with reading books on computers.

References


Note

Send correspondence to James A. Shepperd, University of Florida, Department of Psychology, PO Box 112250, Gainesville, FL 32611-2250; e-mail: shepperd@ufl.edu.
Because the perception of extensive similarity in introductory psychology textbooks is important in the textbook selection process, we examined whether this perception was accurate. We systematically reviewed extant research encompassing 6 textbook dimensions and found that perceived similarity is the product of 2 salient, easily accessible, global text variables: chapter topics and organization. Beyond these variables, textbook homogeneity disappears. We consider stereotyping due to the arduous, complex nature of the textbook selection process as an explanation for this misperception of similarity and urge teachers to consider introductory texts more thoroughly to make the best choices for their individual courses.

Describing the evolution of introductory psychology texts, Weiten and Wight (1992) characterized the 1970s and 1980s as a period of homogenization, attributable primarily to market demands. Clearly, textbook reviews during this period indicated the perception of such homogenization, sporting such titles as “Different Strokes or Reruns?” (Fretz, 1979) and “Are These Books Really Different?” (Jacobs, 1984). Taking a strong stand in his review of four texts, Thomas (1984) asserted that “the difference between texts often does not exceed the famous jnd (just noticeable difference)” (p. 629). This perception of similarity in introductory texts, a common lament of introductory teachers at professional meetings, has even led one e-based company to market a single generic Web-based study guide for all college-level introductory psychology courses (McCollum, 2000). The mere existence of such a study guide provides strong evidence for the perceived homogeneity in introductory texts.

However, this perception of sameness may be an illusion, a misperception that is a product of stereotyping introductory textbooks. Social psychologists have found people are more likely to stereotype in situations that are complex and tax attentional capacity (Kenrick, Neuberg, & Cialdini, 1999). If teachers systematically seek to find the best texts for their courses and students, the process of introductory textbook selection clearly qualifies as complex and arduous. The stereotype that the texts are all pretty much the same would both simplify the selection process and provide a good rationalization for doing so. Because these selections affect 1 to 2 million introductory students annually (Griggs, Jackson, & Napolitano, 1994), it is important to assess whether this perception of similarity is real or illusory for current introductory textbooks. We did so by systematically examining the relevant findings of recent introductory textbook studies.

We hypothesized that this perceived homogeneity is the product of stereotyping—overgeneralization from two very salient, easily accessible, global text variables: chapter topics and organization. We expected a high degree of similarity for these two variables, but a low degree for all other text features. To test this hypothesis, we reviewed recent introductory psychology textbook studies that included quantitative comparisons encompassing six objective dimensions: chapter topics, organization, and extent of coverage; core concepts and key terms; pedagogical aids and data graphs; critical-thinking programs; references; and difficulty level. We did not include other objective dimensions, such as text length and number of chapters, for which obvious differences exist (for a review of these differences, see Griggs, Jackson, Christopher, & Marek, 1999).

Findings and Discussion

Chapter Topics, Organization, and Extent of Topic Coverage

To check chapter topics and organization, we examined the latest compendium of introductory psychology textbooks (Jackson, Griggs, Koenig, Christopher, & Marek, 2000), which described and compared all 41 full-length introductory psychology texts copyrighted between 1997 and 2000. As expected, chapter topics and organization are generally standardized. The following sequence illustrates both the standard topics and their prototypical organization: introduction and research methods, biological processes (biopsychology, sensation–perception, and consciousness), developmental psychology, learning and cognitive processes (learning, memory, thought–language, and intelligence), emotion–motivation, clinical and health psychology (personality, disorders, therapies, and health), and social psychology.

Only 2 (both first editions) of the 41 texts offered novel topic coverage and organization. In the remaining 39 (95%), there was little variance. Only the position of developmental psychology varied often. Sometimes it was positioned more toward the middle following learning and cognitive processes rather than near the beginning after biological processes. Al-
though some texts did offer chapters on nontraditional topics (mainly diversity, sex–gender, and applied psychology), such chapters were still uncommon (see Griggs et al., 1999).

Although the chapter topics and their organization were rather standardized, such global similarity does not preclude divergence in extent of topic coverage. Because teachers also vary in their topic emphases, the mesh between teacher–textbook emphasis is an important factor in textbook selection. Jackson et al. (2000) reported a content analysis for the current 41 introductory textbooks, providing the percentage of text devoted to each of the traditional topics. Considerable variability exists. For example, coverage of biopsychology ranges from 5% to 15%, coverage of sensation–perception from 2% to 12%, coverage of developmental psychology from 5% to 14%, and coverage of social psychology from 3% to 17%. Such variability not only argues strongly against the sameness hypothesis but also gives teachers the flexibility to choose a text that best suits their needs and preferences.

Core Concepts and Key Terms

All studies of this text dimension have shown surprisingly little overlap in core concepts and key terms across introductory texts. Examining the glossaries of 10 texts published between 1994 and 1997, Zechmeister and Zechmeister (2000) identified 2,505 unique terms, with virtually one half (49%) appearing in only 1 text, 8% in at least 8 texts, and only 3% in all 10 texts. Examining a much larger sample of 37 textbook glossaries, Proctor (2000) essentially replicated Zechmeister and Zechmeister’s results. For example, Proctor found that 3,796 (56%) of 6,682 terms were in only 1 text, and only 9 (< 1%) were in all 37 texts. Using a page-by-page analysis for key terms, Landrum (1993) identified 2,742 unique concepts in 6 textbooks, with 58.4% in only 1 text and only 4.6% in all 6. Examining subject indexes of 52 books published between 1980 and 1989, Quereshi (1993) identified 3,813 major terms, with 23.3% in only 1 text. Of the remaining 2,926 terms, only 4.8% were in at least 75% of the texts.

These differences may echo a lack of consensus among introductory textbook authors regarding core terminology. Of the top 100 terms selected in Boneau (1990) as indicators of psychological literacy, Zechmeister and Zechmeister (2000) found only 22% in all the text glossaries examined, with 10% failing to appear in even one glossary. Regardless, it is clear that no substantial core of concepts and key terms exists for introductory texts. Because similarity would definitely be expected on this dimension, the observed dissimilarity provides strong evidence for our hypothesis.

Pedagogical Aids and Data Graphs

Examining all 37 full-length introductory psychology textbooks published between 1995 and 1997 in conjunction with a student survey concerning use of pedagogical aids, Marek, Griggs, and Christopher (1999) quantified differences in the presence of 15 pedagogical aids, finding a range of 4 to 10. Students were most appreciative of those devices most directly related to test preparation (see also Weiten, Guadagno, & Beck, 1996). Although most texts included two highly valued aids—boldface type (100%) and chapter summaries (84%)—inclusion of other highly valued aids—running glossaries (54%), chapter glossaries (16%), end-of-chapter review exercises (11%), and self-tests (8%)—diverged considerably. The presence of less valued aids promoting more structured or elaborative study also varied greatly. For example, chapter outlines, discussion questions, and learning objectives appeared in 100%, 70%, and 14% of the texts, respectively.

In a sample of 11 current introductory texts, Peden and Hausmann (2000) recently found similar variability in the presence and types of data graphs. The total number of data graphs varied from 10 to 144 for their text sample, and the distributions for types of data graph (e.g., line and bar) also revealed considerable variability among texts. Paired with the findings for pedagogical aids, these results strongly indicate lack of similarity across introductory texts in pedagogical assistance.

Critical-Thinking Programs

Examining all 37 full-length introductory psychology textbooks published between 1995 and 1997, Griggs, Jackson, Marek, and Christopher (1998) found that 68% of the texts described critical thinking, with the length of these descriptions ranging from 84 to almost 4,200 words. Texts also differed in the presence and types of special features targeted at promoting critical thinking. Thirty-eight percent featured topical discussions, whereas 8% employed a multistep model of the critical-thinking process. Although Griggs et al.’s assessment did not encompass more subjective evaluation of integrated critical-thinking coverage, it seems likely that approaches to and extent of such integration would vary greatly across introductory texts.

Reference Citations

As with core concepts and key terms, the overlap of reference citations across textbooks is surprisingly minimal. Analyzing 24 introductory psychology textbooks published between 1985 and 1989, Gorenclo and McConnell (1991) found 37,590 unique bibliographic items. However, not one journal article was cited in all texts. Only 3 articles were cited in at least 90% of the texts, 14 in at least 75%, and 51 in at least 50%. These results clearly indicate that introductory psychology textbook authors use rather idiosyncratic criteria for deciding which articles most effectively illuminate specific topics. In brief, the lack of commonality in reference citations between introductory textbooks offers further disconfirmation of the sameness hypothesis.

Level of Difficulty

Considering opinions of textbook authors and editors, Griggs and Jackson (1989) distinguished three levels of text difficulty (high, middle, and low), which reflected breadth and depth of coverage, ties to the experimental literature, the number of pedagogical aids, illustrative programs, and cover-
age of nontraditional topics. Griggs (1999) enlarged this three-level system into a five-level classification scheme by dividing the middle level into three levels (high middle, middle, and low middle). Combining the pedagogical aid and nontraditional chapter variables with a measure of author eminence, he made objective classifications of level for the 37 introductory texts published between 1995 and 1997. These level judgments were then validated via a comparison with independent, experientially-based level judgments by Whitford (1996). The fit was excellent. Although Griggs’s classification procedure is only a first attempt to objectively quantify level, these data clearly indicate that the texts vary greatly with respect to level—one of the most important dimensions of textbook selection.

Introductory textbook readability studies that were more common in the 1970s and 1980s (e.g., Quereshi & Buchkoski, 1979) showed similar findings. Although readability programs (e.g., Flesch formulas) were definitely controversial, the findings of the readability studies, like Griggs’s (1999) results for difficulty level, indicated substantial variability across texts. Thus, the results for both types of studies clearly indicate that introductory textbooks do indeed vary with respect to level of difficulty and readability, at least for all of the procedures that have been used thus far to assess them.

Conclusions

As hypothesized, only chapter topics and organization are generally homogenized. Extent of topic coverage clearly varies, as does level of text difficulty. In addition, texts contain a heterogeneous assortment of pedagogical and critical-thinking programs, with core terminology and references showing minimal overlap. Thus, we conclude that the perception of sameness is indeed illusory. Substantial variance exists between introductory texts on all other dimensions that have been examined, most of which are both relevant and important to text selection. Consequently, we strongly urge teachers not to be misled and think of all introductory texts as the same when selecting texts for their introductory courses. Stereotyping introductory textbooks based on their most accessible features is fraught with inherent dangers similar to those linked to stereotyping people. In both cases, individual strengths, weaknesses, and unique personalities are inappropriately ignored. Similar tables of contents do not entail similar textbooks. To make the best possible text selection for their individual course objectives and students’ abilities, we urge teachers to consider the texts thoroughly on multiple dimensions.

References


Note
Send correspondence to Richard A. Griggs, Department of Psychology, PO Box 112250, University of Florida, Gainesville, FL 32611; e-mail: rgriggs@ufl.edu.

256 Teaching of Psychology
Operant Conditioning Concepts in Introductory Psychology Textbooks and Their Companion Web Sites

Jane P. Sheldon
Department of Behavioral Sciences
University of Michigan–Dearborn

Psychology instructors and textbook authors rate operant conditioning as one of the most essential concepts for students to learn, yet textbook writers, as well as students, can fall prey to misconceptions. This study is a content analysis of the presentation of operant conditioning in introductory psychology textbooks and their companion Web sites to discover if these information sources assist student learning or add to confusion. Results indicate that the failure to refer to changes in the likelihood of the organism’s behavior when discussing operant conditioning was extremely common; this problem should be remedied to reduce students' misunderstandings.

One topic in introductory psychology that contains misconceptions and confusion is operant conditioning (DeBell & Harless, 1992; Lamal, 1995; Tauber, 1988). Introductory psychology instructors and textbook authors rate operant conditioning as one of the most essential concepts for students to learn (Boneau, 1990; Landrum, 1993; Zechmeister & Zechmeister, 2000), and experts in the subfield of behavior analysis (i.e., operant and classical conditioning) rate negative reinforcement, positive reinforcement, and punishment among the most important constructs for students to understand (Buskist, Miller, Ecott, & Crutchfield, 1999). However, students often have a difficult time understanding operant conditioning concepts (Lamal, 1995; Tauber, 1988).

Along with tutorials and activities (e.g., Chrisler, 1988; Graham, 1997; Lukas, Marr, & Maple, 1998; Tauber, 1988), introductory textbooks and their companion Web sites are avenues for student learning. As pointed out by others (e.g., Cooke, 1984; Morris, 1985; Todd & Morris, 1983), unfortunately, textbook writers also can fall prey to misconceptions about behavior analysis. For example, Todd and Morris (1983), in their analysis of 40 psychology textbooks, found that authors sometimes defined reinforcement and punishment without taking into account the organism’s subsequent response rate. In other words, textbook authors often failed to explain that an environmental consequence after a behavior affects the likelihood that the organism will produce the same behavior in the future.

Todd and Morris (1983) conducted their textbook analysis during the 1980s; therefore, there is a need for a more up-to-date analysis of the presentation of operant conditioning terminology in psychology texts. Additionally, this analysis is an investigation of basic operant conditioning concepts, rather than an investigation of more general philosophies within the field (e.g., Lamal, 1995; Todd & Morris, 1983). Finally, an important pedagogical change that has occurred since Todd and Morris conducted their research is the presentation of instructional material on Web sites that accompany textbooks. Therefore, in this content analysis of introductory psychology textbooks I also included their companion Web sites to obtain a more complete picture of the material being presented to students.

Method

Procedure

I contacted the primary publishers of introductory textbooks and requested copies of each introductory text they pub-
lished, acquiring a sample of 36 of the 43 requested texts. In addition, I searched the Internet to find the companion Web sites for each textbook and was able to gain access to 33 sites to include in analyses. (See the Appendix for the list of textbooks and their companion Web sites.) Twice when password-protected Web sites accompanied the text and I did not possess a password, I requested and gained access to the sites by contacting a representative from the publishing company.

Coding

**Textbooks.** For each text I read the information pertaining to operant conditioning terminology in the chapter devoted specifically to learning theories. I did not include sections on reinforcement schedules and shaping because these sections built on previously presented material and thus could appear inadequate or problematic when taken out of context.

For the content analysis I investigated four issues. First, I recorded whether the authors made it clear that the intent of a person doing the reinforcing or punishing is not what matters: Instead it is a change in the likelihood of the organism’s behavior that one must use to determine whether and what sort of conditioning has occurred. Second, I recorded whether authors explained that reinforcers and punishers do not produce the same effect under all circumstances but instead are defined by the effect they produce in a specific circumstance. Third, I counted the number of times authors defined reinforcement and punishers without taking into account how the amount or likelihood of the behavior changed when followed by a consequence. Fourth, I recorded how often the authors presented contradictory, inaccurate, or confusing information (e.g., an appropriate definition with an inadequate or incorrect example, ambiguous wording).

**Companion Web sites.** For each companion Web site I looked through the relevant student learning material (excluding sections concerning shaping and reinforcement schedules) and analyzed the information by focusing on the same four issues that I investigated in the textbooks. The analyses of Web sites took place between February 9 and February 17, 2001.

With the exception of The Psychology Place™, I did not analyze material presented in outside Web sites linked to the textbook’s companion Web site because those sites were not specifically created to be in conjunction with the text. Because The Psychology Place is a site publishers can modify to relate to a specific textbook, it was relevant to include in the study. I was unable to include in the analyses a Wadsworth/Thomson Learning Web site titled WebTutor 2.0 (that accompanied the five textbooks from that publishing company) due to problems gaining access to the different versions of the site. Two textbooks (i.e., Peterson, 1997; Rubin, Peplau, & Salovey, 1993) did not have companion Web sites at the time of data collection. Tavris and Wade (2001) did not have a publisher-based companion Web site but did include access to The Psychology Place. At the time of data collection, relevant material on the site for Smith, Bem, and Nolen-Hoeksema (2001) was not yet on the Web.

### Results

**Intercoder Reliability**

To establish intercoder reliability, two researchers, trained in the use of the coding scheme, independently coded 12 randomly selected textbooks from the sample of 36. Raters established intercoder reliability (number of agreements divided by the total number of agreements and disagreements) separately for each of the four issues investigated in this study. Intercoder reliability ranged from 76% to 93% for the four issues. The coders resolved discrepancies through discussion.

**Analyses**

**Intentions of the conditioner.** Fewer than half (36%) of the textbooks (see Table 1) and only one Web site (see Appendix, Lahey, 2001) made it clear that the intent of the person trying to do the conditioning does not matter. For exam-

<table>
<thead>
<tr>
<th>Textbook Authors</th>
<th>Intent Does Not Matter</th>
<th>Different Effects in Different Circumstances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baron</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bernstein et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carlson &amp; Buskist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis &amp; Palladino</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feldman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gray</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halonen &amp; Santrock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hockenbury &amp; Hockenbury</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huffman et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kassin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kosslyn &amp; Rosenberg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lahey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lefton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morris &amp; Maistro (10th ed.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morris &amp; Maistro (5th ed.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nairne</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passer &amp; Smith</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peterson</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plotnik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubin et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santrock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sdorow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sternberg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tavris &amp; Wade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uba &amp; Huang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wade &amp; Tavris</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weiten</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Westen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood &amp; Wood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wortman et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zimbardo &amp; Gerrig</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zimbardo et al.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The full references for the textbooks are in the Appendix.
ple, some authors discussed how a teacher’s attempt at punishing disruptive behavior by sending a child to the back of the room may actually prove to be positively reinforcing for the child, thus increasing the likelihood of the child’s unruly behavior.

Different effects under different circumstances. Slightly over half (56%) of the texts (see Table 1), but only one Web site (Passer & Smith, 2001), made it clear that reinforcers and punishers do not produce the same effects under all circumstances. For example, authors pointed out that a stimulus (e.g., licorice) that may be reinforcing to some individuals may be punishing to others and that a stimulus that was once reinforcing to an organism (e.g., water) may lose its reinforcing capability at a different time (e.g., immediately after drinking four glasses of water).

Changes in behavior. Thirty-three texts (92%) failed to remind readers at least once that it is a change in the likelihood or amount of a behavior that defines whether an occurrence is reinforcement or punishment, whereas only 24 Web sites (73%) exhibited this error of omission (see Table 2).

Errors, contradictions, and confusion. Table 2 shows that 97% of the texts exhibited at least one confusing explanation, contradiction, or error. Totals per textbook ranged from 1 to 9. Only 30% of the Web sites (see Table 2) showed errors, contradictions, or confusing explanations (ranging from 1–13). Examples included (a) stating that sometimes punishment does not cause a behavior to decline (in which case it cannot, by definition, be punishment), (b) defining reinforcement as giving a reward to increase a behavior (even though that situation describes only positive reinforcement), (c) using the terms positive and negative to refer to the pleasantness of a stimulus (rather than or in addition to whether a stimulus is added or removed), (d) defining negative reinforcement as changing (rather than increasing) a behavior to avoid something, and (e) stating that punishment “generally” or “often” results in a decrease in the behavior (when, by definition, it must result in a decrease in the likelihood of the behavior).

Table 2. Frequencies of Operant Conditioning Explanations in Textbooks and Web Sites That Did Not Refer to a Change in the Likelihood of Behavior and That Contained Errors or Confusion

<table>
<thead>
<tr>
<th>Textbook Authors</th>
<th>Did Not Refer to Change in Likelihood of Behavior</th>
<th>Errors or Confusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Textbook</td>
<td>Web Site</td>
</tr>
<tr>
<td>Baron</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Bernstein et al.</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Carlson &amp; Buskist</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Coon</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Davis &amp; Palladino</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Feldman</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Gray</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Halonen &amp; Santrock</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Hockenbury &amp; Hockenbury</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Huffman, Vernoy, &amp; Vernoy</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Kalat</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Kassin</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Kosslyn &amp; Rosenberg</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Lahey</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Lefft</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Morris &amp; Maistro (10th ed.)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Morris &amp; Maistro (5th ed.)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Myers</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Nairne</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Passer &amp; Smith</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Peterson</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Plotnik</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>The Psychology Place™</td>
<td>—</td>
<td>10</td>
</tr>
<tr>
<td>Rubin, Peplau, &amp; Salovey</td>
<td>8</td>
<td>—</td>
</tr>
<tr>
<td>Santrock</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Sdorow</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Smith, Bem, &amp; Nolen-Hoeksema</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>Sternberg</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Tavris &amp; Wade</td>
<td>7</td>
<td>—</td>
</tr>
<tr>
<td>Uba &amp; Huang</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Wade &amp; Tavris</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Weiten</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>Westen</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Wood &amp; Wood</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Wortman, Loftus, &amp; Weaver</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Zimbardo &amp; Gerrig</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Zimbardo, Weber, &amp; Johnson</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. Full references for the textbooks are in the Appendix.
Discussion

The results make it evident that most introductory psychology textbooks and their companion Web sites may add to students’ confusion about operant conditioning concepts and terminology. These findings say nothing about the accuracy of information in other textbook chapters but do point to the need for instructors to assess textbook and Web site content critically.

The failure to refer to a change in the organism’s behavior when discussing operant conditioning was an extremely common problem in textbooks and Web sites. In Web sites this problem tended to appear frequently in multiple-choice quizzes. For example, a test question that asked students what type of reinforcement or punishment occurred when a child was scolded for misbehaving failed to give any information about changes in the child’s subsequent behavior. One might assume that scolding would cause misbehavior to decrease, but that would be speculation. Scolding may have been a form of attention and therefore a consequence increasing the child’s misbehavior. Therefore, the test question did not give enough information for students to answer the question (even though the site gave students a supposed correct answer).

The failure of the authors to refer to changes in the likelihood of the organism’s behavior ties in with other issues investigated in this study. For instance, “gives praise for participating in class discussions” is a problematic example for two reasons other than the lack of reference to a change in the likelihood of future behavior. First, instructors need to remind students that the intent of the teacher giving the praise is not what determines whether punishment or reinforcement has taken place. Praise may indeed be reinforcing, resulting in an increase in discussion participation. However, for some individuals praise is unwanted, embarrassing, or viewed as patronizing; thus, it may serve as a stimulus that decreases (i.e., punishes) one’s behavior of participating in discussions. Although Skinner (1938) would have decreed the use of cognitive and emotional states to explain the connection between a response and its consequence, this example makes it clear to students that assumptions can be problematic. However, the instructor must continually remind students that behavior analysts pay attention to behaviors rather than to internal mental states because the former can be measured objectively but the latter cannot.

Second, if students are truly to understand operant conditioning concepts, instructors must remind them that reinforcers and punishers can vary from one individual to another and from one time to another. As the results show, only a few textbook authors reminded students of this fact and only one Web site made this notion clear. Instead, textbooks and Web sites implied through omission that certain stimuli are always reinforcers or always punishers. Introductory psychology students likely lack the knowledge base for being able to reject such assumptions and false implications; therefore, it is the responsibility of textbook and Web site authors to present operant conditioning material in an accurate, clear, and consistent manner. Web sites, in particular, offer educators a wide variety of ways to present information and to help students practice using the new material to gain deeper understanding.

Fewer Web sites included errors, contradictions, and confusing explanations than did textbooks; however, this finding is misleading because there was less overall information about reinforcement and punishment presented in Web sites than in texts. Web site developers may want to use these results to guide them in the creation of new, more comprehensive materials and resources to help students learn and remember operant conditioning concepts and terminology. In addition, both textbook and Web site authors should use the findings from this study to reduce the presentation of confusing and inaccurate information related to operant conditioning.

References


Appendix

Textbooks and Web Sites Used for Analyses

*Asterisk indicates that the textbook includes access to The Psychology Place™ at http://www.psychplace.com


Psychology Textbooks: Examining Their Accuracy

Faye B. Steuer and K. Whitfield Ham, II
College of Charleston

Sales figures and recollections of psychologists indicate textbooks play a central role in psychology students’ education, yet instructors typically must select texts under time pressure and with incomplete information. Although selection aids are available, none adequately address the accuracy of texts. We describe a technique for sampling textbooks’ content and evaluating accuracy. Preliminary studies indicated multiple errors in texts we examined. We present a proposed taxonomy of such errors. We speculate that errors are more likely when textbook authors use “deductive” rather than “inductive referencing,” but acknowledge that authors frequently are under pressure to use the former. Discipline-wide efforts to assess textbooks’ scholarly quality could facilitate improvements in accuracy, providing fundamental benefits to the discipline.

Textbooks are of indisputable importance in training newcomers to the discipline and in giving psychology away to many students who take only one or a few psychology courses. Estimates place the size of the U.S. introductory psychology market for new and used textbooks at 1.2 to 1.6 million students per year (V. Knight, personal communication, July 6, 2005; S. Scarrazzo, personal communications, July 11, 2005, August 24, 2006; M. Taflinger, personal communication, August 15, 2006; see also Griggs & Jackson, 1998; B. Miller & Gentile, 1998). Informed estimates suggest that annual domestic expenditures on all psychology textbooks reach $160 to $200 million (S. Scarrazzo, personal communication, July 11, 2005). Such figures imply that introductory textbooks in combination with those used in upper level undergraduate courses constitute a substantial—and potentially massive—part of the educational experience of most psychology students.

In addition to the sheer numbers of textbooks in use, many seasoned psychologists look back fondly on memorable texts as having encouraged their interest in the field and having helped form their ways of thinking about psychological research and theory (see Conkle, 2006). Kihlstrom (2006) surveyed three professional listservs in psychology and received some 130 nominations for textbooks that psychologists remembered as having a positive impact on their professional development.

Textbooks as a genre, however, evoke mixed reactions. On the negative side are views like that described by Morawski (1992), who noted introductory textbooks’ reputation for “deception and inaccuracies, made for the sake of clarity, simplicity, or profit” (p. 162). Weiten and Wight (1992) agreed that introductory texts are sometimes “the object of derision” (p. 487). The equivocal reputation of textbooks may be in part a vestige of an academic value system that for much of the last century ranked teaching and textbook writing below research endeavors (see Boyer, 1990; Halpern et al., 1998; Tyson-Bernstein, 1989; Weiten & Wight, 1992). Such views exist in combination with several published articles that have questioned textbooks’ coverage of certain common topics (discussed subsequently). Collectively, these practices and opinions have supported the view that psychology textbooks are often problematic as scholarly documents.

Such misgivings notwithstanding, a number of observers have suggested that textbooks are generally praiseworthy. There is rather common agreement that textbooks can provide a solid, comprehensive view of the discipline (Christopher, Marek, Dobbins, & Jones, 2004; Griggs & Jackson, 1996; Griggs & Proctor, 2002; Webb, 1991; Weiten & Wight, 1992). Henry and Deka (2004) expressed the belief that instructors can keep up with current scientific literature by assigning and reading new editions of textbooks. A related opinion holds that textbooks are a major influence on students and therefore of appreciable importance (Griggs, 1999; Griggs & Proctor, 2002; Quereshi, 1981). Textbook author David Myers has advised that teachers may “allow the text to cover the discipline” in order to focus on
selected major points during class (H. L. Miller, 2005, p. 278). McKeachie (2002), in his classic book on college teaching, wrote, “With a well-chosen textbook, you may rely upon the students to obtain the basic content and structure of the subject matter through reading” (p. 13). As Myers’s and McKeachie’s statements make clear, an instructor’s choice of textbook can have a decisive impact on the quality of students’ educational experiences. Textbook decisions—wise or unwise—may ultimately have discipline-wide repercussions by influencing students’ approaches to psychological information and even their eventual career choices.

Evaluating Textbooks Is Problematic

Evaluating and selecting textbooks can be time consuming and difficult for several reasons. First, for many undergraduate psychology courses, there exist a large number of available texts (see Christopher, Griggs, & Hagans, 2000; Griggs, Bujak-Johnson, & Proctor, 2004; Marek & Griggs, 2001). Second, in addition to choosing among many texts, instructors must evaluate many features within or accompanying those texts. Relevant features may include difficulty level, breadth and depth of coverage, theoretical issues, accuracy, timeliness, pedagogical aids, and available ancillary materials. Third, the length of the books further complicates the process. Since the 1950s and 1960s, psychology textbooks have tended to be encyclopedic (Weiten & Wight, 1992), a quality that is almost certain to slow the evaluation process. Fourth, the task of deciding on a textbook often takes place just as a semester or quarter is gaining momentum and making its greatest demands on busy faculty members. Fifth, the overall pace of the process is accelerated by the rapidity with which publishers revise and reissue textbooks—typically every 3 years for large-market courses such as introductory psychology. Complex selection decisions may founder if made without due consideration, and yet the enormity and timing of the task can readily overwhelm human decision-making capacity. An instructor teaching three different courses per academic year conceivably might have to select a new text for at least one of those courses every year, while revising another course to accommodate another new text selected the previous semester or year.

It is unrealistic to expect that instructors have time to carefully scrutinize all textbooks available for use in each of their courses. The situation is untenable under the best of circumstances, but particularly so for less-experienced faculty who are still feeling their way through textbook offerings and trying to mesh texts with their developing teaching strategies. Clearly, faculty, students, and the discipline will benefit if aids are available that expedite optimal text selections. Some such aids exist.

Existing Aids for Textbook Selection

Publishers’ representatives supply information that may be useful in choosing texts, but instructors must evaluate such input in light of its marketing objective. The opinions of colleagues, although unlikely to be biased by commercial considerations, may not be optimally suited to the preferences of the instructor making the decision. The scientific literature is another relevant source of information—one with the strength of objectivity to recommend it. Three types of publications may be helpful.

Descriptive and comparative studies compare available textbooks on one or more descriptive dimensions such as length, number of chapters, organization, coverage of core concepts, pedagogical aids, difficulty level, number of references cited, and other features (see Christopher et al., 2000; Griggs, Bujak-Johnson, & Proctor, 2004; Griggs & Marek, 2001; Griggs, Jackson, Proctor, & Cook, 2004; Marek & Griggs, 2001; Griggs, Proctor, & Cook, 2004; Marek & Griggs, 2001; Zecheimer & Zecheimer, 2000). Researchers have also surveyed students regarding the pedagogical aids they find most useful in their textbooks (Marek, Griggs, & Christopher, 2000; Weiten, Deguara, Rehmke, & Sewell, 1999; Weiten, Guadagno, & Beck, 1996). Descriptive and comparative aids may be of real help to instructors or selection committees in reducing a set of possible textbooks to manageable size.

Criticisms of specific content have appeared in the literature and can provide guidance to instructors concerned about accuracy. Such articles include comments on coverage of Wundt and Titchener (Zehr, 2000; see also Burton, 2001), Watson (Benjamin, Whitaker, Ramsey, & Zeve, 2007), Pavlov’s apparatus (Goodwin, 1991), operant conditioning concepts (Sheldon, 2002), recovered memories (Letourneau & Lewis, 1999), Carl Rogers’s theory (Ford & Maas, 1989), and diversity issues (Hogben & Waterman, 1997). Critics also have questioned presentations of schizophrenogenic parenting (Wahl, 1989), ethical
standards for research with children (Ernst, 2003), animal research (Eaton & Sleigh, 2002), and gender issues related to moral development (Hurd & Brabeck, 1997). These articles focus on instances of misinformation that may become widely disseminated as part of what Todd and Morris (1992) called “academic folklore” (p. 1441). Unfortunately, psychology textbooks can be an efficient means for transmitting erroneous folklore.

Textbook reviews appearing in journals such as PsyCRITIQUES provide an additional source of information. Reviewers have typically given a textbook a close reading and can comment on writing style, coherence of information across chapters, or errors that authors and editors may have missed in the process of producing the book (see Landrum, 2000; Hall, 1996). Reviews, more than the other selection aids, consider a book as an integrated document. Instructors often write them, potentially enriching their reviews with their classroom experiences. Reviews are imperfect sources of information, however, in part because not all textbooks are reviewed. The large number of texts published yearly precludes reviewing all of them and leaves would-be adopters with only partial information. Further, reviews represent the opinion of a single observer—a small sample on which to base such an important decision. An additional drawback is that reviews are typically published after a textbook has appeared (Griggs, 1999). Given rapid textbook revision schedules, reviews may come too late to be of practical value in selection decisions, although should a revision not differ much from its previous edition, a review may be informative.

Another Possibility: Evaluating the Accuracy of Textbooks

We think the field of scientific psychology would be well served by ready accessibility of information on the scholarly accuracy of textbooks. We refer especially to the validity of textbooks’ links to the empirical studies they cite. (Although one might expect existing processes of reviewing and editing that occur during production to address this issue, preliminary explorations have revealed that this situation is often not the case.) Providing widely available, sound, useful knowledge about the accuracy of psychology textbooks on a discipline-wide basis would require a coordinated and significant investment of time and work. Here we suggest a relatively simple first step in that direction.

Recent years have witnessed renewed acceptance of the belief that integrating disciplinary knowledge—as in a textbook—is a legitimate scholarly activity (Boyer, 1990; Halpern et al., 1998). “Scholarship” is an abstract concept and, as such, challenging to define denotationally. However, a serviceable definition appears in the work of Diamond (1993), who reviewed definitions from several academic disciplines and concluded that scholarship, in its most general form, involves applying disciplinary expertise in a replicable and documentable way so as to further disciplinary knowledge. Moreover, scholarship should hold up under peer review and have some demonstrable significance. We would add that in the science of psychology, integrative scholarship should also reflect accuracy and allegiance to empirical evidence. Surely instructors should expect scholarship of that caliber in college textbooks.

In this spirit we have examined the accuracy of several introductory-level developmental psychology textbooks. Our approach has been to randomly select passages within a particular book, obtain copies of all references cited in the selected passages, and compare coverage in the textbook to the scientific work that formed its purported base. Over the course of several semesters, we have developed a technique for conducting this type of evaluation.

A Technique for Examining Accuracy

The process we suggest can be time consuming, so we have learned to limit the passages selected to a realistic number. The particular textbook and the number of people working at the task both influence the number of passages we have been able to examine. We find it efficient to work in pairs—an arrangement that minimizes diffusion of responsibility while keeping the paper flow of books and journal articles under control.

Examining all cited references in any textbook would be an inefficient and prohibitively lengthy process; instead, we use a sampling procedure. We first make a random selection of the textbook chapters we will sample. This procedure assures that all chapters have an equal probability of inclusion in our assessment. Ordinarily we have selected about half of the chapters in a particular book. Next we examine the organization of the text to determine how many levels of headings typically appear within chapters. All texts studied to date have relied primarily on two heading levels. Our second step has been to randomly choose one passage, under a second-level heading, in each preselected chapter. We then examine all material...
within each chosen passage. These procedures have most commonly identified eight or nine passages that may vary in length from a few paragraphs to a few pages. Each passage has the coherence of being a complete organizational unit—defined by its heading—in the book. (Representative examples include a section on divorce under a main heading on marriage and divorce and a section on teratogens under a main heading on prenatal environmental influences.) Focusing on passages under second-level headings tends to identify substantive portions of a textbook, rather than introductory vignettes or chapter summaries. We have not included boxed material in our analyses. Our assumption is that accurate scholarship in several randomly chosen passages augurs well for a book generally. Conversely, numerous errors in the samples would raise concerns about a text’s overall scholarly accuracy.

We read all the identified passages carefully, noting all references cited in each. Next, using the text’s reference list, we attempt to obtain copies of all references cited. As we collect references, we read them and compare them to what the text says about them—a task that is time consuming but often interesting in unexpected ways.

We have been surprised to find an alarming number of questionable practices and errors of various types as we have examined the sampled passages of several textbooks. No book has been error free, and all examined so far have contained multiple inaccuracies. This state of affairs necessitated our devising a taxonomy of errors so that we could readily categorize them as we compare references to texts. As shown in Table 1, the list of questionable practices and errors we have encountered—arranged roughly in descending order of seriousness—runs from plagiarism through misrepresentations of primary sources all the way to relatively minor citation errors. It is an empirically derived set of categories in that we have encountered all of them in our examinations of randomly chosen text passages. We present the list in Table 1 as a working document to which users should add if new types of errors come to light. Additional experience with the categories may also warrant collapsing across some of them, removing some, or devising a weighting system to reflect the seriousness of different errors.

As we have used our categorization scheme, we have tested the reliability of our judgments. Using the criterion of both identifying an error and assigning it to an identical category, we have achieved percentage agreement reliabilities of 90% for two independent observers. All disagreements involved categorization, rather than identification, of an error.

Most members of our discipline may assume that rigorous peer review assures the scholarly accuracy of textbooks. However, our explorations call that assumption into question. We have found mainstream introductory-level developmental psychology texts with many errors—some serious—in the passages we sampled. Table 2 shows the incidence of four representative types of errors in five textbooks for which we examined the contents. Because we were still working on our error taxonomy as we examined these texts, we are unable to report on all the types of errors eventually included on the list. We have not identified the texts by name or author, as our purpose here is simply to alert readers to the need for concern. Major publishers produced all five textbooks. To provide an idea of the nature of errors found, a specific example of each type appears in Table 3.

As we engaged in the process of examining texts, we came to discern two fundamental approaches to referencing. One of these, we think, is more likely to give rise to errors than is the other.

### Inductive and Deductive Referencing

**Inductive referencing** is the name we give to a procedure in which an author familiarizes himself or herself with the scientific literature and then crafts a description of it. The story being told grows directly out of the scientific saga described in primary sources. Integrative scholarship at its best is of this type—accurate, empirically grounded, wide-ranging, capable of discovering unexpected trends, and infused with the inherent fascinations of science. **Deductive referencing**, on the other hand, occurs when an author crafts an ostensible review of the literature and seeks references, after the fact, that justify the statements he or she has made. The enterprise then becomes more a matter of defending than of discovering statements about scientific truth.

As we worked our way through the hundreds of sources we read and comparisons we made for this project, it became clear to us that authors frequently relied on deductive rather than inductive referencing. Too often, these deductive references were only loosely related to conclusions advanced in the text. Sometimes they were not related at all.

The pressures that have led to this state of affairs are complex, and we encourage readers not to attribute them glibly to the shortcomings of textbook authors. As noted earlier, textbooks for the most popular...
Table 1. Definitions of Questionable Practices and Citation Errors

<table>
<thead>
<tr>
<th>Practice or Error</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plagiarism*</td>
<td>- Quoting another author without enclosing the quoted passage in quotation marks, with or without providing a reference.</td>
</tr>
<tr>
<td></td>
<td>- Copying material from a source word-for-word with a few words changed, with or without providing a reference.</td>
</tr>
<tr>
<td></td>
<td>- Borrowing the observations, analyses, arguments, or findings of another author without giving credit to that author (i.e., using the ideas of another author without citing or clearly acknowledging that author's original work).</td>
</tr>
<tr>
<td>“Paraphragiarism”&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>- Extended paraphrases of others’ words; one-to-one idea correspondences; structural similarities; substituting another's original words with comparable, synonymous language (usually more disguised and difficult to detect than instances of plagiarism).</td>
</tr>
<tr>
<td></td>
<td>- Some phrases and sentences echo the source in such a way that the sentence structure, content, themes, or examples are virtually identical.</td>
</tr>
<tr>
<td>Misrepresenting a fact or finding reported in a cited reference</td>
<td>- Includes misunderstanding, overgeneralizing, exaggerating, understating, or otherwise obscuring the original author’s meaning (i.e., offering an inaccurate representation of the empirical observations, results, statistical outcomes, or conclusions presented in the source).</td>
</tr>
<tr>
<td></td>
<td>- Only clear and obvious examples of such practices are to be scored (based on explicit, directly observable evidence).</td>
</tr>
<tr>
<td>Misrepresenting a theoretical concept or argument presented in a cited reference</td>
<td>- Includes misunderstanding, overgeneralizing, exaggerating, understating, or otherwise obscuring the original author’s meaning (i.e., conveying a distorted or fabricated interpretation of the cited reference by misconstruing the author's underlying assumptions, hypotheses, point of view, abstractions, or inferences, as portrayed in the referenced work).</td>
</tr>
<tr>
<td></td>
<td>- Only clear and obvious examples of such practices are to be scored (based on explicit, directly observable evidence).</td>
</tr>
<tr>
<td>Inappropriately or unnecessarily citing a reference</td>
<td>- Citing a reference that is unrelated to the topic or issue presented (i.e., providing a reference that has no purpose or relevance).</td>
</tr>
<tr>
<td></td>
<td>- Presenting a broad, general statement and subsequently providing an unnecessary text citation, introducing a source that only loosely and vaguely applies to the topic being discussed in the text.</td>
</tr>
<tr>
<td>Failing to report a relevant key or central point made in a reference cited</td>
<td>- Referring to a relevant source but missing a main point that should have been included (i.e., insufficiently extracting and summarizing the source's primary message or leading argument).</td>
</tr>
<tr>
<td></td>
<td>- Omitting or denying that a particular finding, result, or piece of information occurred when it actually did occur.</td>
</tr>
<tr>
<td></td>
<td>- To be scored only when the justification for doing so is clear and obvious (based on explicit, directly observable evidence).</td>
</tr>
<tr>
<td>Citing a secondary (or tertiary) source as if it were a primary source</td>
<td>- Citing the author of a secondary (or tertiary) source as if that author were the originator of the fact, finding, or concept described.</td>
</tr>
<tr>
<td>Citing a secondary (or tertiary) source when a primary source would have been available and more appropriate</td>
<td>- Noting the secondary (or tertiary) nature of the source cited but failing to cite the original source of a fact, finding, or concept when the original work described would have been available.</td>
</tr>
<tr>
<td></td>
<td>- To be scored only when the original source is widely and readily available.</td>
</tr>
<tr>
<td>Citing a reference that is functionally impossible to find</td>
<td>- Citing a reference that could not be located and retrieved despite several attempts to track down the reference (i.e., reporting information from a source that was unattainable, even though all possible means to find the cited work were used).</td>
</tr>
<tr>
<td></td>
<td>- Omitting publication details needed to retrieve a source and therefore making it virtually impossible to identify the source without resorting to inference or guesswork.</td>
</tr>
<tr>
<td>Reporting a fact, finding, or theoretical idea that is functionally impossible to find in the cited reference</td>
<td>- Reporting a fact, finding, or idea that does not appear in the table of contents or index of the cited work.</td>
</tr>
<tr>
<td>Practice or Error</td>
<td>Definition</td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Presenting information that cannot be located in the cited reference, despite three meticulous and orderly attempts to discover it. (Three thorough and systematic searches may minimize potential careless mistakes and ensure that the evaluator has not overlooked the key information.) Attempts include reading the content of the source closely with an eye for detail or, if the source is a book with articles by several authors, checking to see if the relevant material is contained in another author's article.</td>
<td>– Presenting information that cannot be located in the cited reference, despite three meticulous and orderly attempts to discover it. (Three thorough and systematic searches may minimize potential careless mistakes and ensure that the evaluator has not overlooked the key information.) Attempts include reading the content of the source closely with an eye for detail or, if the source is a book with articles by several authors, checking to see if the relevant material is contained in another author's article.</td>
</tr>
<tr>
<td>Usually applies to book-length references.</td>
<td>– Usually applies to book-length references.</td>
</tr>
<tr>
<td>Inappropriately citing an unpublished article as a reference</td>
<td>– Citing a conference presentation that is 7 or more years old (based on the textbook's year of publication).</td>
</tr>
<tr>
<td>– Citing an unpublished source when a published (and more applicable) version existed before the textbook was published.</td>
<td>– Citing an unpublished source when a published (and more applicable) version existed before the textbook was published.</td>
</tr>
<tr>
<td>– Does not apply to personal communications or to “classic” articles that seem to be available only in unpublished form.</td>
<td>– Does not apply to personal communications or to “classic” articles that seem to be available only in unpublished form.</td>
</tr>
<tr>
<td>Citing a reference that is clearly out of date</td>
<td>– Citing a book that has since undergone revision, with the revised edition being as relevant (or more relevant) than the cited reference.</td>
</tr>
<tr>
<td>– Presenting information from a considerably older and dated source as if it were up-to-date and current.</td>
<td>– Presenting information from a considerably older and dated source as if it were up-to-date and current.</td>
</tr>
<tr>
<td>– Reporting outdated coverage of factual information from an ongoing area of research that is continuously and frequently being updated.</td>
<td>– Reporting outdated coverage of factual information from an ongoing area of research that is continuously and frequently being updated.</td>
</tr>
<tr>
<td>– Only clear and obvious examples of this practice are to be scored (based on explicit, directly observable evidence).</td>
<td>– Only clear and obvious examples of this practice are to be scored (based on explicit, directly observable evidence).</td>
</tr>
<tr>
<td>Citation error: Author’s name</td>
<td>– Error in spelling of author’s first name, initials, or surname, either in a text citation or in the reference list.</td>
</tr>
<tr>
<td>– Incorrect order of authors’ names, either in a text citation or in the reference list.</td>
<td>– Incorrect order of authors’ names, either in a text citation or in the reference list.</td>
</tr>
<tr>
<td>– Failure to cite one or more contributing authors, either in a text citation or in the reference list. (Does not apply to references with more than six authors.)</td>
<td>– Failure to cite one or more contributing authors, either in a text citation or in the reference list. (Does not apply to references with more than six authors.)</td>
</tr>
<tr>
<td>Citation error: Journal, periodical, or book title</td>
<td>– Error in listing of the title of the journal, periodical, or book where the cited reference is to be found.</td>
</tr>
<tr>
<td>– Includes omission of a subtitle</td>
<td>– Includes omission of a subtitle.</td>
</tr>
<tr>
<td>Citation error: Article title</td>
<td>– Error in the title of an article.</td>
</tr>
<tr>
<td>– Includes omission of a subtitle.</td>
<td>– Includes omission of a subtitle.</td>
</tr>
<tr>
<td>Citation error: Publication date</td>
<td>– Date error in a text citation, in the reference list, or both.</td>
</tr>
<tr>
<td>Citation error: Numbering of volume, issue, or edition of reference</td>
<td>– Volume, issue, or edition error in text citation, in the reference list, or both.</td>
</tr>
<tr>
<td>Citation error: Page numbering</td>
<td>– Page-number error in a text citation, in the reference list, or both.</td>
</tr>
<tr>
<td>Citation error: Alphabetizing errors in reference list</td>
<td>– Failure to provide inclusive page numbers of a chapter in an edited book.</td>
</tr>
<tr>
<td>Citation error: Alphabetizing errors in reference list</td>
<td>– References not listed in proper alphabetical order in the reference list.</td>
</tr>
<tr>
<td>Citation error: Publishing data</td>
<td>– Errors in identifying a publisher (e.g., failure to provide location of a conference proceeding, to correctly give the name and Web address of an electronic reference, or to spell publisher's name correctly).</td>
</tr>
<tr>
<td>– Not to include older books that may have been reissued by multiple publishers.</td>
<td>– Not to include older books that may have been reissued by multiple publishers.</td>
</tr>
<tr>
<td>Citation error: Other</td>
<td>– Any citation error not described previously.</td>
</tr>
</tbody>
</table>

Note. The questionable practices and citation errors described are arranged roughly according to perceived seriousness, from most serious (i.e., plagiarism) to least serious (i.e., citation error: other).

<sup>a</sup> It is possible for both plagiarism and “paraphragism” to occur in a single instance (e.g., in a single passage citing a single reference). <sup>b</sup> Word suggested by Levin (1992, p. 12).

Psychology courses are typically revised every 3 years. Publishers market their texts as up-to-date, timely, and containing references to the most recent scientific findings. Such pressures regarding time and referencing essentially dictate that authors revising their texts will have to rely heavily on deductive referencing. Time would rarely, if ever, permit a new synthesis of an entire field every 3 years. Although the first
Table 2. Errors per Reference Citations Examined in Five Developmental Psychology Textbooks

<table>
<thead>
<tr>
<th>Textbook</th>
<th>Misrepresenting a Fact or Finding</th>
<th>Misrepresenting a Theoretical Concept or Argument</th>
<th>Inappropriately or Unnecessarily Citing a Reference</th>
<th>Citation Error (All Types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11/25 (44.0%)</td>
<td>0/25 (0%)</td>
<td>4/25 (16.0%)</td>
<td>7/25 (28.0%)</td>
</tr>
<tr>
<td>B</td>
<td>12/36 (33.3%)</td>
<td>1/36 (2.7%)</td>
<td>9/36 (25.0%)</td>
<td>13/36 (36.1%)</td>
</tr>
<tr>
<td>C</td>
<td>3/16 (18.7%)</td>
<td>0/16 (0%)</td>
<td>5/16 (31.2%)</td>
<td>5/16 (31.2%)</td>
</tr>
<tr>
<td>D</td>
<td>5/19 (26.3%)</td>
<td>0/19 (0%)</td>
<td>4/19 (21.0%)</td>
<td>4/19 (21.0%)</td>
</tr>
<tr>
<td>E</td>
<td>18/64 (28.1%)</td>
<td>2/64 (3.1%)</td>
<td>3/64 (4.6%)</td>
<td>31/64 (48.4%)</td>
</tr>
</tbody>
</table>

Discovering Textbooks’ Accuracy

Although we believe it will ultimately prove desirable to mount a discipline-wide effort to assess the scientific veracity of textbooks, at the present time, the only realistic way to discover such information is on one’s own. One practical approach is for instructors to enlist the help of independent study students and use the technique we have described to examine textbooks they are already using. The procedure creates an interesting and doable one- or two-semester independent study project in which both faculty member and student can learn a great deal—not just about the texts they are evaluating but also, as a bonus, about the primary literature they will read. By the time of a text’s next revision, an instructor should have a better sense of whether the revision is likely to meet his or her expectations regarding accuracy.

It will be even more advantageous, we believe, if psychology—as an academic discipline—devises a way collectively to monitor and report on the accuracy of the textbooks its students read. Systematic information of this type would be a conspicuously useful additional aid for textbook selection. By arising from within the discipline, the information should remain unbiased by marketing pressures. Such an undertaking would, of course, require time, money, and human capital (in the form of organizational skill coupled with an absolute sense of objectivity, fairness, and accuracy). However, we believe it would yield enduring benefits across the discipline. Perhaps a cooperative effort will emerge naturally if individual instructors become more familiar with the empirically based accuracy of the textbooks they assign.

Table 3. Examples of Errors Found in Texts Examined

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misrepresenting a fact or finding</td>
<td>Textbook says older adults who engaged in life review showed significant improvement in depressive symptoms. Reference reports significant improvements in depressive symptoms for all groups studied—the life-review group and two control groups—so improvement cannot be attributed to life review.</td>
</tr>
<tr>
<td>Misrepresenting a theoretical concept or argument</td>
<td>Textbook states that many education experts say children learn best through concrete experiences. Reference focuses on electronically mediated, interactive learning and does not address the text’s point. However, many relevant references are available to support the text’s theoretical statement.</td>
</tr>
<tr>
<td>Inappropriately or unnecessarily citing a reference</td>
<td>Textbook reports several facts, figures, and dates regarding the AIDS epidemic. None of this information appears in the reference cited.</td>
</tr>
<tr>
<td>Citation error (all types)</td>
<td>Reference in text does not appear in the textbook’s “References” section.</td>
</tr>
</tbody>
</table>
Concluding Thoughts

Given the demonstrable importance of textbooks in educating the next generation of psychologists, it behooves members of our discipline—particularly the many who teach—to assure that texts used by students are of unimpeachable scholarly and scientific quality. We propose a technique for sampling a text’s content and evaluating its accuracy. In developing and using the technique, we have constructed a system for categorizing scientific and scholarly inaccuracies, which we offer to readers as a starting point for their explorations of scholarly quality. We encourage such explorations with the strong suggestion that colleagues also attempt to discover creative, efficient, accurate, and fair ways to share knowledge about this important matter. We believe that pressures toward more solid scholarship in textbooks, arising from within our discipline, can produce fundamental benefits for psychology and psychologists.

References


Notes

1. We thank Ludy T. Benjamin, Jr. and Michael M. Marcell for helpful comments on the article and Mona Stribling, Leisa Morrill, Amanda Floyd, Heather Matusz, and Joy Wray, who participated in earlier phases of this project.

2. Send correspondence to Faye B. Steuer, Department of Psychology, College of Charleston, Charleston, SC 29424; e-mail: steuerf@cofc.edu.
The Representation of Applied Psychology Areas in Introductory Psychology Textbooks

Charlotte W. Haselhuhn and Kerri L. Clopton
Department of Educational Psychology and Foundations, University of Northern Iowa

Many psychology majors indicate helping others as a reason for majoring in psychology, yet many enter positions not closely related to the field. This discrepancy may be due to a lack of student knowledge of the applied areas of psychology. The purpose of this study was to investigate the coverage of clinical, counseling, industrial/organizational, and school psychology in a sample of current, full-length introductory psychology textbooks. This study explores the extent and accuracy of coverage of applied areas in 32 full-length textbooks published between 2001 and 2006. Industrial/organizational psychology had the most extensive coverage, followed by clinical, counseling, and school psychology. Counseling psychology had the lowest rate of accuracy in coverage.

The U.S. Bureau of Labor Statistics (2006) predicted that employment of psychologists will grow faster than average through 2014 because of an increasing need in applied areas. Many undergraduate psychology majors appear to be interested in applied subfields, as they report entering the field because of a desire to help others (Metzner, Rajecki, & Lauer, 1994). However, a minority of psychology majors go on to graduate study and only a small percentage enter closely related professions after earning their bachelor's degree (Borden & Rajecki, 2000; Finno, Salazar, Frincke, Pate, & Kohout, 2006; Tsapogas, 2004). It is possible that few psychology graduates enter applied areas because of a lack of knowledge. For example, undergraduate psychology students reported the greatest knowledge about clinical and counseling psychology when compared to other applied areas, yet their self-reported knowledge level was between “very little” and “moderately familiar” (Stark-Wroblewski, Wiggins, & Ryan, 2006). The underrepresentation of applied areas in textbooks may explain this lack of knowledge (Dixon, Vrochopoulos, & Burton, 1997; Griggs, Jackson, & Napolitano, 1994). School psychology, a field suggested by the U.S. Bureau of Labor Statistics (2006) to have the best job opportunities, is covered less thoroughly than other applied areas (Dixon et al., 1997; Wise, 1981).

Prior investigations of coverage of applied fields have focused on single specializations (Maynard, Geberth, & Joseph, 2002; Wise, 1981) or a comparison of two specializations (Lucas, Blazek, Raley, & Washington, 2005), or have sampled only a subset of introductory texts (Dixon et al., 1997). It is difficult to compare the results of previous studies concerning coverage because of different methods and materials. Furthermore, little is known about the accuracy of the existing coverage of the applied areas of psychology. This study investigated the extent, as well as the accuracy, of introductory textbook coverage of the applied areas of psychology.

Method

Sample

We obtained textbook titles and authors from the Office of Teaching Resources in Psychology (OTRP) compendium of introductory psychology textbooks (Koenig, Daly, Griggs, Marek, & Christopher, 2004; available online at http://www.lemoyne.edu/OTRP/introtexs.html) and requested the most recent edition of the 36 full-length textbooks listed, along with any other introductory texts, from the local representative of the publishing companies. One representative did not respond to our request, two texts were out of print, and one newer version we received did not appear to be a general introductory textbook. A full-length text not on the compendium was sent by one representative and was included in the study,
resulting in a final sample of 32 full-length textbooks with copyright dates from 2001 through 2006. Twenty-three of these texts were more recent editions of those listed on the OTRP compendium.

Procedure

Three advanced graduate students in school psychology and two faculty researchers collected information from the textbooks. Information from each textbook was initially collected by two readers, and a third reader was involved when initial agreement was not adequate. Readers evaluated each text for coverage of the applied psychology areas, including clinical, counseling, industrial/organizational (I/O), and school psychology, by determining whether the applied psychology area was present in the glossary or index and by counting the number of words within the text that referred to the subject area. We examined pages listed in the index for each area as well as chapters focused on different areas and careers in psychology. Readers counted the total number of words referring to the particular area that appeared in text, tables, or figures. Readers also counted words in appendices, when appropriate. The count included words in sentences that contained the name of the particular area or clearly referred to the area. We evaluated the accuracy of coverage by comparing the content to a description of the profession available through the related divisions of the American Psychological Association (APA; Society for Industrial and Organizational Psychology, Inc., n.d.; Society of Clinical Psychology, n.d.; Society of Counseling Psychology, 1999) and the National Association of School Psychologists (NASP, 2003). We used the NASP description of school psychology because we could not locate a description on the APA Division 16 Web site at the onset of the study. We considered a text’s description to be accurate if it did not contradict information contained in the description provided by the area’s respective APA division or, in the case of school psychology, by NASP. It was possible for a description to be accurate but incomplete.

Initial agreement on inclusion of an area in the text, index, or glossary of a textbook was substantial, with kappa values of .64 to perfect agreement. A kappa value above .61 is considered substantial (Landis & Koch, 1977). Readers reached the criterion of 97% agreement on initial word counts 63% of the time. We resolved differences between readers by obtaining a third opinion, discussing the content, or, in the case of number of words, recounting the content until we met a 97% level of agreement.

Results

Extent of Coverage

We considered an area of psychology covered if the text specifically mentioned the area anywhere in the book, whether in the index, glossary, text, tables, or figures. In each section of the textbooks (text, index, and glossary) the pattern of the extent of coverage was the same: Clinical psychology received coverage in the most textbooks, followed closely by I/O psychology. School psychology received coverage in the fewest texts (see Table 1). The mean number of words addressing clinical, counseling, I/O, and school psychology differed significantly, $F(3, 93) = 11.35, p < .001$. Pairwise comparisons with least significant difference adjustment for multiple comparisons indicated the mean number of words for each area differed from the mean number of words for all other areas. I/O psychology received the most coverage in terms of total words, followed by clinical, counseling, and school psychology. There was also considerable variation in the extent of coverage within each area of psychology. I/O coverage showed by far the greatest range, with two textbooks providing no word coverage and one textbook providing 10,206 words (see Table 1).

Accuracy of Coverage

All of the textbook descriptions of clinical and I/O psychology were accurate, and 22 of the 24 descriptions of school psychology (91.7%) were accurate. The accuracy of counseling psychology definitions was surprisingly low, with only 9 of 29 (31.0%) textbooks including accurate descriptions when compared with the description provided by Division 17 of the APA (Society of Counseling Psychology, 1999). The majority of inaccuracies found for counseling psychology were statements that counseling psychologists work with clients with less serious “everyday life” problems, implying or stating that they do not work with those who have serious psychiatric and personal difficulties. We considered such statements inaccurate because the description provided by Division 17 specifically states that counseling psychology is “unique in its attention to both normal developmental issues and to problems associated with physical, emotional, and mental
Table 1. Representation of Applied Psychology Areas in Introductory Psychology Textbooks

<table>
<thead>
<tr>
<th>Area</th>
<th>Clinical</th>
<th>Counseling</th>
<th>I/O</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Represented in book</td>
<td>32</td>
<td>100.00</td>
<td>29</td>
<td>90.63</td>
</tr>
<tr>
<td>Index</td>
<td>31</td>
<td>96.88</td>
<td>28</td>
<td>87.50</td>
</tr>
<tr>
<td>Glossary</td>
<td>16</td>
<td>50.00</td>
<td>10</td>
<td>31.32</td>
</tr>
<tr>
<td>Description accuracy&lt;sup&gt;b&lt;/sup&gt;</td>
<td>32</td>
<td>100.00</td>
<td>9</td>
<td>31.03</td>
</tr>
<tr>
<td>Mean number of words (SD)</td>
<td>309.33 (252.53)</td>
<td>139.78 (94.03)</td>
<td>2324.08 (3640.55)</td>
<td>71.73 (71.62)</td>
</tr>
</tbody>
</table>

*Note.* Results are reported for the sample of 32 textbooks unless otherwise indicated.
<sup>a</sup> Ten of the 26 textbooks that included school psychology described it as part of the field of educational psychology.
<sup>b</sup> Percent accurate is based on the number of textbooks that included a description of the specialty area: clinical, n = 32; counseling, n = 29; I/O, n = 30; school, n = 24.

Discussion

The results of this study suggest that the information in introductory texts regarding careers in applied areas of psychology is limited. The lack of introductory textbook coverage of school psychology is particularly striking. Stark-Wroblewski et al. (2006) found that undergraduate psychology majors reported lower levels of familiarity and were less interested in a career in school psychology than in counseling or clinical psychology. The lack of interest in school psychology careers may be due to a lack of, or inaccurate, knowledge. Although 92% of the textbooks that included descriptions of school psychology did so accurately, the low number of words indicated a lack of detail about the field.

The difference between I/O psychology and school psychology in extent of coverage in the current study is striking. Projections by APA (2007) and the U.S. Bureau of Labor Statistics (2006) indicate that job opportunities will expand considerably for both areas in the near future, yet the average number of words addressing I/O psychology in the textbooks surveyed was 32 times the average number of words addressing school psychology. Textbook authors and publishers may deem this level of I/O coverage appropriate given the growing demand in the field, yet school psychology had the smallest amount of coverage, although experiencing a national shortage in practitioners (Curtis, Hunley, & Grier, 2004) and having the best job prospects of the applied areas of psychology (APA, 2007). Although not necessarily related to textbook coverage, it is noteworthy that the Society for Industrial and Organizational Psychology (2006) has made it very convenient for instructors to include I/O psychology in their courses. This division of APA provides a set of 14 teaching modules in areas of I/O psychology, complete with PowerPoint slides, supplemental materials, and additional resources (see http://www.siop.org/Instruc/InGuide.aspx).

Limitations

We must acknowledge the limitations of this study. First, we were unable to obtain a complete sample of introductory texts. Although we obtained 94% of the available full-length introductory psychology textbooks, a lack of response from one publisher’s representative resulted in the absence of two texts from the sample. Second, we found the development of specific rules for actual word counting and determining accuracy of content that covered a broad range of circumstances very difficult. Our discussion of these issues led to clearer, but not comprehensive, guidelines due to the extensiveness of some content.

Recommendations

Psychology faculty, especially those who teach introductory courses or advise undergraduate students, need to remain current on the job outlook and opportunities in the various fields of psychology. The APA publishes brochures and books on careers in psychology and provides a significant amount of information on its Web site. Faculty and students should explore other professional organizations that represent the applied disorders” (Society of Counseling Psychology, 1999, p. 1). Accuracy of coverage for each of the four applied psychology areas appears in Table 1.
areas of psychology to supplement the information in textbooks. For example, the NASP Web site (www.nasponline.org) provides extensive information on the roles and functions of school psychologists, the job outlook, and graduate programs. OTRP (Society for the Teaching of Psychology, 2007) maintains a Web site designed to assist career advisors in psychology, containing helpful information about the applied fields (see http://teachpsych.org/otrp/resources/resources.php?category=Advising). The U.S. Bureau of Labor Statistics Occupational Outlook Handbook is also available online (www.bls.gov/oco/ocos056.htm) and provides information on careers, salary, educational requirements, and job outlook. Faculty also should consider inviting applied psychologists from the community into their classes or other departmental events for students.

References


Appendix: References for the 32 Textbooks Used in This Study


Notes

1. We express our appreciation to Kelly Beck, Erin McCoy, and Sarah Wakefield for their assistance in this research.
2. The data from this study were presented at the National Association of School Psychologists 2006 Conference, Anaheim, CA.
3. Send correspondence to Charlotte W. Haselhuhn, 617 Schindler Education Center, University of Northern Iowa, Cedar Falls, IA 50614–0607; e-mail: charlotte.haselhuhn@uni.edu.
Extra Credit Exercise: A Painless Pop Quiz

B. Michael Thorne
Mississippi State University

The extra credit exercise (ECE)—a nonpunitive pop quiz—is a useful adjunct to the traditional lecture-style course. The ECE potentially encourages class attendance, fosters preclass preparation, gives students (and instructors) feedback on their learning (and teaching) of the course material, provides students with test-type questions, perhaps reduces test anxiety, and gives students extra credit toward their final point totals. In addition, students often find ECEs a desirable feature of the course, something that is almost never true of the traditional pop quiz.

Motivating student preparedness for each class is one of teaching’s greatest challenges. In a course with regularly scheduled examinations, students often file their study and preparation to a few days immediately preceding the test, interspersed with periods of inactivity. That is, students frequently respond to scheduled examinations as though they were on a fixed-interval schedule, with little responding (studying) immediately after the examination and cramming immediately before the test (e.g., Mawhinney, Bostow, Laws, Blumenfeld, & Hopkins, 1971). Operant conditioning research indicates that the way to achieve more steady responding is to provide reinforcement randomly, on a variable interval schedule, for example. The pop quiz is a traditional method of testing at unannounced, random intervals.

Although a major purpose of pop quizzes is to encourage preclass preparation, often students are punished for being unprepared. This punitive aspect and the unpredictability of pop quizzes undoubtedly contribute to students’ distaste for them (e.g., Combs, 1976). Illustrating their entrenched aversion, desensitization training to reduce test anxiety completely removed negative feelings for multiple-choice, essay, and mathematics exams in 40 test-anxious undergraduates but failed to remove such anxiety completely for pop quizzes and timed tests (Crouse, Deffenbacher, & Frost, 1985).

However, pop quizzes can be useful. For example, Kerkman, Kellison, Piñon, Schmidt, and Lewis (1994) found that students who wrote test items for pop quizzes performed significantly better on the tests than students who did not write questions. Additionally, Bell (1997) found the use of anonymous pop quizzes (students were not identified) worthwhile for identifying problem areas in student understanding of the material presented. Over the years, I have developed a pop quiz variation combining the benefits of frequent, unannounced testing with features designed to make the quizzes less aversive, which I call the extra credit exercise (ECE).

At the beginning of the semester, I tell students that I will be giving brief exams at irregular intervals. These exams will be solely for extra credit and will not count against the student, no matter how poorly he or she performs. Generally, each ECE is worth up to one extra point; partial credit is possible, with the credit from all the ECEs summed and rounded to a whole number added to the student’s point total after the final grade distribution is decided. Theoretically, a student can make an “A” for “A” work on the major exams, even if he or she has acquired no ECE points. In practice, this rarely happens, as students with few or no ECE points have generally missed many classes, and there is a strong class attendance-grade correlation in my courses. In the class in which I initiated ECEs—introductory psychological statistics—the top students will generally earn between 10 and 15 extra points from the quizzes (from a maximum of 17 possible), which amounts to fewer than 4% of the total points from the hour exams (400).

In statistics, most ECEs consist of one problem (e.g., computing the Pearson $r$, performing a one-sample $t$ test). For a completely correct problem, the student receives one extra point, with fractions of one point deducted for errors such as making the wrong significance decision, recording an incorrect critical value from a table, or making a careless arithmetical error. In other courses (e.g., introductory psychology, history of psychology, physiological psychology), ECEs consist of multiple-choice, matching, fill-in-the-blank, and short-answer essay questions, with points assigned according to the nature of the question. On an ECE, points correct divided by the total number of points produce a fractional total such as 0.5 or 0.2. The length of time for the ECEs depends on the class; in statistics, the quiz may take 30 min or more (e.g., for one-way ANOVA), whereas in other classes, the time is usually fewer than 10 min. In my classes, I generally average one ECE a week, with ECEs in statistics given following the discussion of a particular technique (e.g., one-way ANOVA, $t$ test for independent samples).

Evaluation of the Assignment

In addition to the extra credit, which can make the difference between one letter grade and the next, ECEs have several benefits. For example, graded and returned promptly, ECEs illustrate questions or problems students can expect on the major tests, thus becoming important adjuncts to the students’ pretest study material. Additionally, ECEs encourage both class attendance and preclass preparation. Because I
collect the papers and immediately go over the correct answer or answers, students quickly learn the effectiveness of their study practices and whether they understood the material tested. They are then in a position to ask questions about problem areas. Furthermore, these problem areas are highlighted for me as I grade the ECEs. For example, if I find that many students in my statistics class do not know how to draw conclusions from a significance test, I can emphasize this point in my next lecture.

The sheer number of ECEs should help desensitize test-phobic students. With tests at monthly or longer intervals, there is a lot of pressure to do well on them, and some students find this pressure debilitating (Crouse et al., 1985). Smaller, more frequent quizzes should help defuse this test anxiety. Also, frequently given quizzes, discussed in class and then graded and returned, illustrate that the instructor is interested in students learning the material. For some students, the knowledge that their instructor is committed to teaching them something can be a powerful motivator.

Despite all they have going for them, ECEs are primarily just pop quizzes with a name change. Do students like them any better than more traditional pop quizzes? One of the questions on our faculty evaluation is “What did you like best about this course?” Forty of 162 students who responded to the question in seven different classes (24.7%) indicated they particularly liked the ECEs, with one student writing, “[I like] extra credit exercises because you get feedback on how well you are progressing without being punished.”

In summary, ECEs at least partially reduce the aversion associated with traditional pop quizzes, encourage class attendance and preclass preparation, give students examples of test questions, and enable students to earn extra credit toward their final grades. In addition, the exercises provide feedback to the students on their level of understanding and acquisition of material in the course at the same time that they give the instructor feedback on the success of his or her teaching.

References


Note

Send correspondence to B. Michael Thorne, Drawer 6161, Mississippi State, MS 39762; e-mail: bmt2@ra.msstate.edu.

Undergraduate Student Perceptions of the Impact of Faculty Activities on Education

Christopher A. Bjornsen
Longwood College

Psychology and nonpsychology undergraduate students (n = 212) provided their perceptions of the impact of professors’ out-of-class responsibilities on student education. Students perceived professor activities directly related to instruction and 1-to-1 contact as most important to the quality of education and placed substantially less value on research and service activities. Results of factor analysis suggested an alternate model of student perceptions. Resolving the discrepancies between the 2 models—as well as the needs of students, faculty, and administrators—involves bridging the perceived gap between the importance of teaching, research, and service.

Responding to concerns over accountability, and in an effort to improve the quality of education, faculty and administrators in higher education have increasingly employed student assessments of faculty effectiveness. Recent summaries of the literature on student evaluations of teaching (SET) have argued that SET are reliably associated with a variety of student traits, class content, and workload factors (Marsh & Roche, 1997) as well as instructor skills such as instructional delivery, ability to facilitate interactions, and evaluation of student learning (d’Apollonia & Abrami, 1997). Data also support the impact of soliciting student feedback concerning instructor weaknesses in the context of the classroom on teaching performance (Perlman & McCann, 1998). Although concerns regarding the proper administration, interpretation, and use of SET remain, the ability of student evaluations to provide useful assessments of professor abilities in the classroom is largely supported (McKeachie, 1997).

If students can provide meaningful assessments of faculty effectiveness, they may also be capable of providing valuable feedback regarding the effect of professors’ out-of-class activities on student learning. This area does not appear to have received research attention, a notable deficit given the time and energy that faculty expend in various activities outside of the classroom. This study was conducted at a medium-size, state-supported liberal arts institution where faculty are evaluated based on their effectiveness in three well-defined areas: quality of teaching, professional activity, and college and community service. It seemed logical, then, to assess the degree to which students perceive that professor activities in these three areas affect the quality of student education.

Method

Participants

Students enrolled in seven different classes participated in the study (n = 212, mean age = 20.29 years). Within the sample, there were 48 (22.6%) psychology majors and 164
Faculty Forum


Available online: 20 Nov 2009


To link to this article: http://dx.doi.org/10.1207/S15328023TOP2703_09

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Extra Credit: Gifts for the Gifted?

Marjorie S. Hardy
Eckerd College

Faculty often discuss whether to offer extra-credit opportunities to students. I describe the statistical advantages (or disadvantages) of allowing students to complete extra credit. Results indicate that only better students take advantage of optional extra-credit assignments, suggesting the need to reconsider their usefulness.

A request heard frequently during the semester, particularly as it is drawing to a close, is “Can I do some extra credit to bring up my grade?” Often, the students making this request have not been performing well throughout the semester, and instead of pacing themselves consistently and diligently over the past 14 to 15 weeks, have decided to “sprint” in the homestretch. To avoid this dilemma, I have regularly instituted a semester-long extra-credit option for all students. In this article, I present the details of this option and its differential effect on students’ grades.

Previously, researchers have investigated instructors’ use of and opinions about extra credit (e.g., Hill, Palladino, & Eison, 1993; Norcross, Dooley, & Stevenson, 1993) and have found that the use of extra credit is particularly prevalent in the social sciences (Norcross, Horrocks, & Stevenson, 1989). Few, however, have addressed the issue of whether extra credit helps those who need it most. Junn (1995) found that a skills-based extra-credit assignment helped marginal students pass, but she also found that the extra credit was particularly prevalent in the social sciences (Norcross, Horrocks, & Stevenson, 1989). Few, however, have addressed the issue of whether extra credit helps those who need it most.

When extra credit is optional, however, who chooses to complete it? Previous studies on the volunteer research participant may shed some light on this issue. Rosenthal and Rosnow (1975) described one type of volunteer as the “good student” who does not need the extra credit to obtain a good grade. On the other hand, Henley and Savage (1994) found no significant differences in grades between students who did and did not volunteer in optional research studies outside of the classroom. However, they did find significant positive correlations between the students’ scores on two tests and the total amount of extra credit earned.

I conducted this study to answer the question, “Who completes extra credit?” Is it the marginal student who needs the extra points to pass, or is it the good student who works diligently throughout the semester, taking every opportunity (including time to study) to make a good grade? Based on Rosenthal and Rosnow’s (1975) description of the good student who completes extra credit, conversations with colleagues, and anecdotal observations, I hypothesized that the students who would choose to complete extra credit would be those who least needed it.

Method

In two sections of introductory psychology, undergraduate students (N = 54) took four noncumulative exams, spaced evenly throughout the semester. They also completed three laboratory assignments during the semester.

Consistent with Palladino, Hill, and Norcross’s (1999) suggestions that extra-credit opportunities be equally accessible to all students, pedagogically sound, built into the course structure, and comprise less than 5% of the final grade, I offered extra credit to all students at the beginning of the semester in the following manner. Students could improve their test grades throughout the semester by summarizing up to five articles (scholarly or popular) related to the topics we were covering for the upcoming exam. Students turned in copies of the articles and one-page summaries (explaining the relation of the articles to course topics) on the days of each exam. The students could earn up to 5 points on each exam, 1 point for each article they correctly summarized. If they correctly summarized the maximum 5 articles per exam, they could therefore earn up to 20 points of extra credit (5 points per exam, four exams), which in turn could conceivably account for a 3.3% improvement in their final course grade.

Results

Only about half the students (n = 26) chose to take advantage of the extra-credit option, and only 12 of those students turned in extra-credit articles for every exam. There was no difference in participation rate between the two sections, X^2(1, N = 54) = 3.60, p > .05. Of the 26 students who turned in summaries, the average number of (correctly completed) summaries across the four exams was 2.66 (SD = 1.48); only two students correctly completed five articles for each exam. Correctly completed summaries related the material in class to the topic of the article; only eight summaries were deemed unacceptable.

Completing the extra-credit assignments resulted in an average increase of 2.66 points (SD = 1.48) in final grades for those students who chose to complete at least one assignment. The difference in the grades of these students, before and after adding the extra credit, was statistically significant, t(25) = –9.15, p < .001, d = .65. Practically speaking, however, the additional points resulted in a letter grade improvement for only six students, with three students moving from a “B” to an “A” and three students moving from a “C” to a “B.” (Other students saw practical improvements in terms pluses and minuses.)

An independent samples t test compared the preextra-credit grades of those students who chose to complete the assignments and those who did not. The results were significant, t(52) = 2.38, p < .02, with the students who completed extra credit having a higher preextra-credit average (M = 84.25, SD = 7.32) than the students who did not complete the assignments (M = 78.96, SD = 8.87).
Discussion

The results of this study support the hypothesis that the students who chose to complete optional extra-credit assignments were those students who needed them the least. Approximately half of the students in two sections of introductory psychology completed some of the assignments, with few taking full advantage of the opportunity. Completing the assignments resulted in statistically higher averages overall, but in actual grade changes (e.g., from a “B” to an “A”) for only a few students. One limitation of this study was the relatively small sample size. Instructors currently using extra credit may want to collect data to assess the relative effectiveness of the opportunity for their high- and low-achieving students.

So, should an instructor offer extra credit to students? Students certainly seem to perceive the opportunity as potentially beneficial to their grade, particularly if the assignment is worded with respect to the number of total points they can potentially earn. In a subsequent survey of 87 students in three introductory psychology classes, I described the extra-credit assignment and asked two questions: (a) How likely is it that you would complete extra-credit summaries if you improve your grade by up to 20 points (added to your exam total); and (b) How likely is it that you would complete extra-credit summaries if you could improve your grade by up to 3.3%? (Both options, counterbalanced across the questionnaires, yield the same number of additional points to the final course grade.) Participants responded along a 5-point Likert scale ranging from 1 (not at all likely to complete the assignments) to 5 (very likely to complete the assignments). I also asked them to rate the impact that they believed the extra credit would have on their final grade, to report their current grade point average (GPA), and to estimate their expected grade in introductory psychology. The results were as I hypothesized. For the 20-point option, the mean likelihood rating was a 4.14, with 50.6% of students believing the extra credit would significantly improve their final grade. For the 3.3% option, the mean likelihood rating was a 3.72, with only 25.3% of students believing the extra credit would significantly improve their final grade. Both the likelihood rating differences and impact differences were statistically significant, $t(86) = 3.79$ and 3.80, respectively ($p < .001$ for both). An interesting finding was that GPA and expected grade in the class were not significantly correlated with the items, suggesting that what students think they would do and what they might actually do may be very different. In fact, of the students who responded to the survey, 87.5% reported that they would be at least somewhat likely to complete the 20-point assignments (75.1% for the 3.3% option), percentages significantly greater than the 48% in my classes that actually completed at least one assignment.

There are other advantages to the use of extra credit, as well. One advantage of semester-long, optional extra credit, available to all students, is that it makes it more difficult for the last minute student to ask, “Is there something extra I can do?” (It at least makes it easier for the instructor to say, “No,” pointing out that the student has not taken advantage of such opportunities all semester.) Moreover, students who diligently apply themselves all semester and who take advantage of the extra-credit opportunities are unlikely to ask for the chance to raise their grade even higher because they already have high averages in the first place.

Thus, if the purpose of extra credit is to help marginal students pass, this purpose is not likely to be met with optional assignments. If the only students who choose to complete and therefore benefit from optional extra-credit assignments are those who need them the least, then instructors may want to reconsider the ways in which they make such assignments (if they make them at all).

References


Notes

1. An earlier version of this article was presented at the 23rd Annual National Institute on the Teaching of Psychology, St. Petersburg Beach, FL, January 2001.
2. I thank Randolph A. Smith and three anonymous reviewers for helpful comments on an earlier draft of this article.
3. Send correspondence to Marjorie Hardy, Department of Psychology, Eckerd College, 4200 54th Avenue South, St. Petersburg, FL 33711; e-mail: hardyms@eckerd.edu.

Notetaking Predictors of Test Performance

Robert L. Williams
Alan Eggert
Department of Educational Psychology
University of Tennessee

In this study, we determined the relation between full-semester notetaking and test performance variables in a large undergraduate human development course. Although the notetaking framework equally weighted readings and lectures, notetaking over lectures (particularly accurate notetaking) proved more promising in accounting for test performance measures than did notetaking over reading materials.

A recent review by Armbruster (2000) highlighted the facets and formats of lecture notetaking that contribute most
We propose a procedure for increasing student participation, particularly in large classes. The procedure establishes a token economy in which students earn tokens for participation and then exchange those tokens for extra credit. We evaluated the effectiveness of the procedure by recording the degree of participation in an introductory psychology class before, during, and after implementation of the token economy. Results revealed that the amount of directed and nondirected participation increased during the token economy and returned to baseline after removal of the token economy. Furthermore, students responded faster to questions from the instructor during the token economy than during baseline, and this decrease in response latency continued even after removal of the token economy.

A considerable literature attests to the importance of active learning in which students engage and process course material rather than passively receive it (e.g., Benjamin, 1991; Bligh, 2000; Bonwell & Eison, 1991). One way instructors can facilitate active learning is to challenge the class periodically with relevant questions and encourage students to offer questions and comments. However, instructors may avoid this form of classroom interaction because of a phenomenon we call “the silence,” the uncomfortable time following the instructor’s question when no one responds. The silence is a particular problem in large classes in which students feel relatively anonymous and are reluctant to participate (McKeachie, 2002). Instructors can use a variety of techniques to combat the silence, such as waiting out the silence (Kendall, 1994), calling on students by name (Gurung, 2002), or initiating small group discussions (McKeachie, 2002). In this article, we present another method for breaking the silence that is effective and easy to use, particularly in large classes.

Our method relies on extra credit to reinforce participation. Other faculty have used extra credit as an incentive to improve exam performance (Junn, 1995; Nation & Bourgeois, 1978), read journal articles (Carkenord, 1994), seek writing assistance (Oleyn, 1992), demonstrate critical thinking (Junn, 1994), improve behavior modification projects (Barton, 1982), and avoid procrastination (Lloyd & Zylla, 1981; Powers, Edwards, & Hoehle, 1973). Our method creates a token economy in which students earn tokens for participation. Immediately following participation, the instructor presents a token to the student. At the end of class, students exchange their tokens for extra credit toward their course grades.

Hodge and Nelson (1991) also used reinforcement to shape classroom participation. In their study, the instructor wrote students’ initials on the board and placed plus marks next to the initials of students who exhibited the desired amount of participation. Although similar to our method, Hodge and Nelson’s procedure differs from ours in several ways. For instance, their procedure is feasible only in small classes, whereas our method is relatively easy to use in classes of almost any size. Indeed, the first author has successfully used our method in classrooms that seat as many as 200 students. Also, Hodge and Nelson evaluated the effectiveness of their technique based on students’ self-reported participation. In contrast, we evaluated the effectiveness of our method more objectively by having a research assistant observe the degree of student participation prior, during, and after the token economy.

Method

Participants

Sixty-three undergraduate students enrolled in an introductory psychology course at the University of Central Arkansas participated in the study.

Procedure

The class met 75 min twice weekly for 16 weeks. We conducted the study over the final 11 class meetings of the term. During each of these 11 class meetings, the instructor periodically directed relevant questions to the class, and students who wanted to answer the questions raised their hands. The instructor then called on students in the order in which they raised their hands until a student answered the question correctly. If no one raised a hand within 60 sec following a question, the instructor announced the answer and continued with the lecture.

The first 4 of the 11 class meetings served as the baseline period. During this time, students did not receive any explicit reward for answering a question correctly. Over the next 4 class meetings, the instructor implemented the token economy. The instructor announced that the first person to answer a question correctly would receive a token. The tokens were wooden checker pieces purchased from a local hobby store. The pieces were heavy enough to throw, but light
enough not to cause injury if they missed their target. At the end of each class meeting, students could exchange each token for one point added to their next exam grade. Each exam point was worth 0.25% of the course grade. If students did not turn in their tokens at the end of the class meeting, those tokens were void, and students could not exchange them for extra credit in the future. This rule ensured that the instructor had to keep a supply of tokens for only one class meeting and avoided claims of lost tokens. During the final 3 class meetings, the instructor discontinued the token economy and informed students that they could no longer earn tokens for correct answers. As required by our university’s institutional review board, the instructor also provided students who had not earned extra credit during the token economy with alternative extra credit opportunities during the removal period. After the removal period, the instructor fully debriefed students about the study.

During each of the final 11 class meetings, a research assistant sat in the last row of the classroom where she had an unobstructed view of all students and posed as a student in the class (e.g., by pretending to take notes). The research assistant recorded the amount of directed participation (number of students who raised their hands in response to a question from the instructor), latency to participation (amount of time following each question until the first hand was raised), and amount of nondirected participation (number of times any student spontaneously asked the instructor a question or engaged the instructor in discussion). The research assistant measured latency using a hand-operated digital stopwatch, which she kept hidden at all times.

Results

The instructor asked 16 questions during baseline, 14 during the token economy, and 16 during removal. Overall, the instructor asked a mean of 4.18 questions per class meeting. Only once did no student raise a hand following a question from the instructor. We recorded and analyzed this question, which occurred during baseline, as zero directed participation, but removed it from the analysis of latency to participation. Table 1 presents a summary of all three dependent measures across the three phases.

### Table 1. Means for the Dependent Measures Across the Three Phases

<table>
<thead>
<tr>
<th>Dependent Measure</th>
<th>Baseline</th>
<th>Token Economy</th>
<th>Removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed participation/question</td>
<td>1.63&lt;sub&gt;a&lt;/sub&gt;</td>
<td>3.64&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.63&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Latency to participation/question&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.16&lt;sub&gt;a&lt;/sub&gt;</td>
<td>0.56&lt;sub&gt;b&lt;/sub&gt;</td>
<td>2.93&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Nondirected participation/class period</td>
<td>9.50&lt;sub&gt;a&lt;/sub&gt;</td>
<td>21.75&lt;sub&gt;b&lt;/sub&gt;</td>
<td>11.00&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Note. Values within a row not sharing a subscript are significantly different (<i>p</i> ≤ .05).

<sup>a</sup>Time latencies are reported in seconds.

Directed Participation

We analyzed amount of directed participation using focused chi-square tests. We adjusted the expected frequencies to control for the different number of questions across the three phases. Compared to baseline, significantly more students raised their hands in response to the instructor’s questions during the token economy, χ<sup>2</sup>(1, N = 77) = 11.85, <i>p</i> < .001. Furthermore, students raised significantly fewer hands during removal than during the token economy, χ<sup>2</sup>(1, N = 77) = 11.85, <i>p</i> < .001, but the number of hands raised during removal was not significantly different from baseline, χ<sup>2</sup>(1, N = 52) = 0.00.

Latency to Participation

We conducted a one-way ANOVA of the latency data. Each question from the instructor, rather than each student in the class, constituted the unit of analysis. The ANOVA revealed a significant difference between the mean latencies of the three phases, F(2, 42) = 8.23, <i>p</i> = .001, η = .53. Tukey’s honestly significant difference (HSD) test indicated that students raised their hands significantly faster during the token economy than during baseline (<i>p</i> = .001). However, Tukey’s HSD tests showed that latency to participation during removal was not significantly slower than during the token economy (<i>p</i> > .20), but was significantly faster than during baseline (<i>p</i> = .05).

Nondirected Participation

We analyzed amount of nondirected participation using focused chi-square tests. We adjusted the expected frequencies to control for the different number of class meetings across the three phases. Compared to baseline, students spontaneously participated significantly more during the token economy, χ<sup>2</sup>(1, N = 125) = 19.21, <i>p</i> < .001. However, during removal students spontaneously participated significantly less than during the token economy, χ<sup>2</sup>(1, N = 120) = 11.56, <i>p</i> < .001. Furthermore, nondirected participation did not significantly differ between baseline and removal, χ<sup>2</sup>(1, N = 71) = 0.38, <i>p</i> > .44.

Discussion

As we hoped, the amount of directed and nondirected participation dramatically increased following the implementation of the token economy. Students were more than twice as likely to raise their hands following a question during the token economy than during baseline. Likewise, students were more than twice as likely to ask questions and to make comments spontaneously during the token economy than during baseline, even though the instructor did not directly reinforce this form of participation with tokens. Thus, in general, students appeared more willing to contribute to the class during the token economy. Once the instructor removed the to-
We believe the token economy procedure is a simple and effective means of breaking the silence, especially in large classes. In addition, the procedure serves as an excellent demonstration of operant conditioning and the utility of token economies. Indeed, during the removal period, while the instructor described token economies, one student spontaneously noted that the instructor had used a token economy to increase students’ participation. We believe this sudden connection promotes an “a-ha” experience for the class and a deeper understanding of the material. Furthermore, the first author has noticed an increase in student attendance, enthusiasm, and preparation when he has used the token economy. Students have commented that they enjoy the procedure because it makes class more exciting and interactive.

Finally, the token economy system described in this study is flexible and easily adapted to an instructor’s teaching style. We understand that some instructors do not like to use extra credit in their courses. However, instead of extra credit toward the students’ course grades, tokens could be worth credit toward “purchasing” desirable options, such as dropping a quiz or being excused from the final exam (see Komaki, 1975). Alternatively, instructors could replace tokens with other easily delivered rewards, such as candy. As long as students perceive a contingency between some positive reinforcer and their participation, instructors may develop variations to suit their teaching style.

References


Gurung, R. (2002, June). Sleeping students don’t talk (or learn): Enhancing active learning via class participation. In P. Price (Chair), Active learning in the classroom: Overview and methods. Symposium conducted at the 14th annual meeting of the American Psychological Society, New Orleans, LA.


**Notes**

1. We thank Bill Lammers and Timothy Johnston for their helpful comments on an earlier draft of this article.
2. Send correspondence to Kurt A. Boniecki, University of Central Arkansas, Department of Psychology and Counseling, 201 Donaghey Avenue, UCA Box 4915, Conway, AR 72035; e-mail: kurtb@mail.uca.edu.
In this article I address issues associated with online courses, focusing on introductory psychology. Data from 2 studies compared student attrition, performance, and evaluation in classroom sections and online sections. In Study 1, students chose to enroll in online or classroom sections; in Study 2, students were assigned to a section. Data indicate that attrition was similar in the 2 instructional formats. In Study 1, online students scored nonsignificantly lower on course tests and the final exam. Online students were more likely to fail the course. However, online students evaluated the course similarly to the classroom students. In Study 2, performance and evaluation were not different for the online and the classroom sections.

Higher education, like any field, is subject to occasional trends, breakthroughs, and even paradigm shifts. Consider the overhead projector and the enormous impact this simple technology had on classroom presentation. Currently there are two forces with the potential to revolutionize higher education. These trends are increased use of instructional technology and distance learning (Willis, 1994). Many administrators believe increased use of technology and distance learning will help institutions attract students and use faculty resources most effectively. Hence, increased use of instructional technology and distance learning are often institutional goals (Council for Higher Education Accreditation, 1999).

Although instructional technology and distance learning can take different forms, they come together in the online course. In a fully online course, the entire course consists of Web pages that students visit regularly. These pages communicate course content, deliver assignments, serve as a forum for class discussions, and may even be the site of exams. The faculty and students may never meet face to face. There are numerous potential benefits of online instruction. First, online courses allow institutions to serve nontraditional students. Students may complete their coursework at their leisure, thus allowing students who cannot attend conventional classes to have access to higher education. Instructors may also find themselves freed of the requirement of teaching class at the same time every day. As a matter of fact, students and faculty need not be at the same campus or even the same institution. For example, I taught an online course for Athens Technical College (ATC) while teaching at the University of Innsbruck in Austria one summer.

Although there is still little research on the issue, it appears that the addition of online courses can increase an institution’s overall enrollment. Ridley, Bailey, Davies, Hash, and Varner (1997) found that the addition of wholly online instruction (eight online courses; one online degree program) increased the total number of full-time-equivalent students at Christopher Newman University (CNU). The authors concluded that online courses reduce the negative effects of distance and scheduling on enrollment.

A second advantage to online courses is their appeal to traditional on-campus students. These students find surfing the Web fun and engaging. They like the idea of taking a course online and may be more involved and motivated (Ridley, 1995; Ridley et al., 1997; Waschull, 1997a).

Third, using the Web for instruction allows instructors to link students to a huge collection of resources. With a click of the mouse, students can read primary sources, view images and illustrations, or visit specialized sites that address a particular topic. For example, in psychology there are many high-quality Web sites available for class use (e.g., Psych Web, Dewey, 1999).

Fourth, the online environment can foster a greater degree of communication and closeness among students and faculty (Knight, Ridley, & Davies, 1998; Waschull, 1997b). Students often feel comfortable asking a question by e-mail that they might have been too hesitant or shy to ask in class. In response, the instructor can take as much time as needed to answer a question without fear of losing precious class time.

Finally, the online learning environment may be more rigorous than the traditional classroom environment. Ridley (1998) compared the learning demands in online and traditional classroom courses. He surveyed students to measure perceptions of time spent studying, writing, reading, thinking, doing library research, and so on. Students enrolled in online courses reported that they spent more time (a) writing and developing writing skills, (b) reading both required and suggested texts, (c) using critical thinking skills and reflecting on course content, (d) using the library for research projects, (e) studying, and (f) discussing course material with peers and instructor.

Although there is still little research on the issue, there are also some legitimate concerns about a fully online course. Kerka (1996) and Waschull (1997b) identified numerous potential problems with the online approach, including reliance on student initiative, limited access to computers in some areas, potential for information overload, risk of social isolation, limited technological resources of the student or the institution, and increased potential for academic dishonesty. In addition, many instructors are also concerned about the quality of instruction of online courses.
Although there has been limited research on the issue of the quality of online courses (and online courses probably vary greatly in quality), research by Ridley and colleagues (Knight et al., 1998; Ridley, 1995, 1998) addressed issues of attrition, performance, and evaluation in online courses at CNU. Ridley (1995) found that online courses at CNU were comparable to classroom courses in rigor, performance, and student satisfaction. However, online courses had a higher withdrawal rate and students reported them to be more intensive and demanding than regular courses. In a follow-up assessment of the online program (Ridley, 1998), a comparison of students’ grade point averages (GPAs) revealed that GPAs for online courses were significantly lower than GPAs for classroom courses (about .03 grade points lower). Knight et al. (1998) also reported that students in online and classroom courses reported similar levels of educational satisfaction and provided similar instructor evaluations.

Although this research paints a relatively favorable picture of online courses, there is a serious restriction to the conclusions one can draw. Students who elect to take online courses may be different from students who do not. If, for example, only better students enroll in online courses, then the fact that online courses appear comparable to traditional classroom courses may be due to the above-average comprehension and performance of these self-selected students. Indeed, Knight et al. (1998) also found that online students had a stronger knowledge base entering their program of study and maintained that advantage during their education (based on an objective measure of students’ knowledge base in their major field of study given to all incoming students at the University of Phoenix). The participants in the Knight et al. study were not only more knowledgeable entering their program; they were also significantly older and more likely to be White and male. The online students also had more work experience, higher salaries, and were more likely to intend to pursue a graduate education. Although demographics differ from one institution to another, the potential for self-selection in online courses is great. This is one of the more difficult issues for research to address.

Another important issue in the assessment of online courses is the comparability of online and classroom-based courses. The two instructional formats are fundamentally different. To determine if student learning is equivalent in the two approaches, it is important to address the comparability of the courses under study and ensure that they are similar in every aspect other than instructional method.

The research reported here addressed these issues in an online introductory psychology course. In particular, I was interested in comparing student attrition, performance, and satisfaction in comparable sections of online versus classroom introductory psychology courses (Study 1). Furthermore, I wanted to control for the effects of self-selection in the online course (Study 2).

Study 1

Method

To ensure that the online course and the classroom course were comparable, I taught one section of introductory psychology online and one section of introductory psychology in the classroom. Participants had the option of enrolling in either of the sections or in sections taught by other faculty.

Participants. A total of 33 students participated in the study. Fourteen students enrolled in the online section and 19 students enrolled in the classroom section. Students in the online course were 71% White, 64% women, and ranged in age from 21 to 47. Students in the classroom section were 79% White, 74% women, and ranged in age from 21 to 48.

Instructional method. The classroom course met five times weekly for 50 min. Course meetings consisted of lecture, supplemented by discussions. Students also completed five written assignments and read assigned chapters from Psychology (Benjamin, Hopkins, & Nation, 1994). To ensure that the classroom students did not access any of the online materials, the materials were password protected and I closely monitored activity on the Web site.

Online students visited the course Web site (Waschull, 1999), read four to five online lectures a week, visited about 10 relevant Web sites a week, completed five written assignments, and read assigned chapters from Psychology (Benjamin et al., 1994). Online students completed all of their coursework, including tests and written assignments, online. All online students had access to the ATC computer labs (located in Athens, GA, and 11 surrounding counties) and computers located in public libraries. Many online students also had access to home computers and computers located in their place of employment.

Although the two instructional formats led to some differences between the sections (e.g., the online students visited relevant Web sites whereas the classroom students engaged in discussions), the course content was identical in the two sections. I based the online lectures on the lecture notes I used in the classroom section of introductory psychology and the content of the online lectures was identical to the content of the classroom lectures. The students in both sections also received the same textbook reading and written assignments.

Measures. To assess student performance, I gave the online and classroom sections four identical tests and an identical, comprehensive final exam. Tests consisted of 30 to 40 multiple-choice items, 8 to 10 fill-in-the-blank items, and one essay question. The final exam consisted of 50 multiple-choice items. To evaluate student satisfaction, participants completed the standard Course and Instructor Evaluation form given to all ATC students.

Results

Attrition. Fourteen students enrolled in the online section; 1 dropped the course. Nineteen students enrolled in the classroom course; 3 dropped. The attrition was 7% in the online section compared to 15% in the classroom section. Fisher’s exact test was nonsignificant ($p = .43$), indicating that the observed frequencies were not significantly different. Chi-square analyses indicated that the distribution of race, sex, and age did not differ for the online and the classroom sections: $\chi^2(2, N = 33) = 2.98, ns; \chi^2(1, N = 33) = .34, ns$;
\( \chi^2(12, N = 33) = 12.96, \text{ns} \), respectively. Furthermore, the distribution of these variables did not differ for those students who withdrew and those who completed the course: race, \( \chi^2(2, N = 33) = 3.21, \text{ns} \); gender, \( \chi^2(1, N = 33) = .06, \text{ns} \); and age, \( \chi^2(12, N = 33) = 12.08, \text{ns} \).

Performance. For each test and the final exam, I used \( t \) tests to compare test scores for the two sections. The means appear in Table 1. Due to the number of comparisons, the traditional significance level of \( p = .05 \) was modified using the Bonferroni correction. The corrected level of significance for five comparisons was \( p = .01 \). Although the mean test scores were slightly lower in the online section for all but one test, this difference did not achieve statistical significance. To determine whether the observed pattern of nonsignificant mean differences translated into different final course grades, I compared the proportion of students who passed (or failed) the course in the two sections. A chi-square analysis revealed that the two courses did differ in pass–fail rates, \( \chi^2(1, N = 27) = 4.44, p < .05 \). Students were significantly more likely to fail in the online section based on the pass–fail rate of students in the classroom section.

Evaluation. Students rated the course and the instructor on a 5-point scale ranging from 1 (never) through 5 (always) on nine dimensions. The mean ratings for each item appear in Table 2. For eight of the nine items, the mean score was slightly higher in the online section than the classroom sections (means were identical on the ninth item). Due to the large number of comparisons, the traditional significance level of \( p = .05 \) was modified using the Bonferroni correction. The corrected level of significance for nine items was \( p = .005 \). The mean difference was not significant for any of the nine evaluation items.

**Discussion**

The results of this study indicate that the online introductory psychology course did not differ significantly from the classroom introductory psychology course in terms of student attrition and satisfaction. The online students’ test performance was also similar to traditional classroom students’ test performance. However, the pattern of nonsignificant mean differences translated into a higher rate of course failure. This pattern is consistent with the Ridley’s (1998) finding that students have somewhat lower GPAs in online versus classroom courses.

Because I taught both the online and classroom sections, I covered the same material in the online lecture notes and in the classroom lectures, and because I gave both sections identical assignments and tests, this study controlled many of the content-related factors that could lead to performance differences between the two groups. However, Study 1 did not address the issue of self-selection; it is possible that these results occurred because of the particular type of student who chose to sign up for the online versus the classroom course.
The goals of Study 2 were to replicate the results of Study 1 with more students and to eliminate the opportunity for self-selection.

Study 2

Method

In the academic quarter following Study 1, I again taught one section of introductory psychology online and one section of introductory psychology in the classroom. Participants had the option of enrolling in either of these sections or in a number of sections taught by other faculty members. However, students were not told that one of the sections I taught would be online. After registration, I selected one section to be the online section. I met the students at the assigned time on the first day of class and explained that the course would be fully online. Although I encouraged students to remain in the online section, they had the opportunity to change sections if they wished. All students agreed to remain in the online section. Other than this important difference, Study 2 was conducted in exactly the same manner as Study 1.

Participants. A total of 41 students participated in Study 2. Eighteen students enrolled in the online section, and 23 enrolled in the classroom section. Students in the online course were 89% White, 89% female, and ranged in age from 21 to 54. Students in the classroom section were 78% White, 48% female, and ranged in age from 20 to 39.

Results

Attrition. Eighteen students enrolled in the online section; 2 dropped the course. Twenty-three students enrolled in the classroom section; 2 dropped. This number reflects 11% attrition in the online section compared to 8.8% in the classroom section. Fisher's exact test of these frequencies was nonsignificant ($p = .60$).

Chi-square analyses indicated that the distribution of race and age did not differ between the online and the classroom sections, $\chi^2(2, N = 41) = 1.19, ns$; $\chi^2(18, N = 41) = 22.73, ns$, respectively. The distribution of sex was significantly different for the online and classroom sections, $\chi^2(1, N = 41) = 7.57, p = .006$; significantly more men enrolled in the classroom section. The distribution of these variables did not differ for students who withdrew and students who completed the course: race, $\chi^2(2, N = 41) = .47, ns$; age, $\chi^2(18, N = 41) = 25.10, ns$; and sex, $\chi^2(1, N = 41) = .55, ns$.

Performance. For each test and the final exam, I used $t$ tests to compare the mean scores. The means appear in Table 1. Due to the number of comparisons, the traditional significance level of $p = .05$ was modified to .01 using the Bonferroni correction. Although the mean test scores were lower in the online section for all but one test, this difference did not achieve statistical significance. These results are similar to those from Study 1.

To determine whether the observed pattern of nonsignificant mean differences translated into different final course grades, I compared the proportion of students who passed (or failed) the course in each section. A chi-square analysis revealed that the two courses did not differ in pass–fail rates, $\chi^2(1, N = 10) = .01, ns$. In Study 2, students were no more likely to fail in the online section compared to the pass–fail rate of students in the classroom section of introductory psychology.

Evaluation. Students in both sections completed the standard Course and Instructor Evaluation form given to all ATC students. The mean ratings for each item appear in Table 2. Due to the large number of comparisons, the traditional significance level of $p = .05$ was modified using the Bonferroni correction. The mean difference was not significant at $p < .005$ for any of the nine evaluation items. Although the mean difference was not significant, it is interesting to note that the pattern of ratings is consistent and opposite that observed in Study 1, when the students selected the online section.

General Discussion

The combined results of Studies 1 and 2 indicate that online and traditional classroom introductory psychology courses are similar on some important measures. This finding was true of students who chose to take their introductory psychology course online as well as for students who were assigned to this form of education. Attrition in the online section was low and comparable to classroom sections whether the online section was chosen or assigned to students. Attrition rates did not differ based on race, age, and sex. Likewise, test performance was not significantly lower in online sections whether the online section was chosen or assigned. Despite the fact that none of the 10 test performance comparisons achieved statistical significance there was a fairly consistent pattern; mean performance was lower in the online section in 8 out of the 10 comparisons. Did this pattern translate into lower overall course grades? The results are mixed: In Study 1, the proportion of students passing the course was significantly lower in the online section. However, in Study 2 the proportion of students passing the course was not significantly different in the online section and the classroom section.

Although the differences did not achieve significance, a noteworthy pattern in student evaluations emerged across Study 1 and Study 2. Students in Study 1 who had chosen the online course rated the course and instructor slightly higher than students in the classroom section on each of nine items. On the other hand, students in Study 2, who were assigned to the online course, gave somewhat lower evaluations than students in the classroom section on each of the nine items.

The usefulness of online instruction for removing barriers to higher education combined with this preliminary evidence that students perform comparably in an online introductory psychology course indicates that online instruction in introductory psychology may be a useful addition to traditional forms of instruction. Particularly, these results indicate that the success of students in online courses is not completely a product of self-selection. When self-selection was controlled,
attrition, performance, and evaluation were similar in online and classroom sections.

It is important to interpret the results of these two studies cautiously. Because of the large number of mean comparisons performed, I used a modified significance level to protect against Type I error. Some differences that would have achieved significance at a lower level of significance are not significant using the modified level. Furthermore, the relatively small number of participants in the two studies limits generalizations. Finally, there are still many important questions about online instruction that must be answered. In particular, research addressing issues such as retention of information and performance in subsequent psychology courses is sorely needed before online instruction in psychology becomes widespread.

References

Waschull, S. B. (1997b, September). Teaching and learning over the WWW. Paper presented at the meeting of the National Alliance of Community and Technical Colleges, Biloxi, MS.

Note

Send correspondence to Stefanie B. Waschull, Athens Technical College, 800 Highway 29 North, Athens, GA 60601; e-mail: waschul@aati.edu.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
Pedagogical research on Web-based learning and instruction has not kept pace with the proliferation of Web-based courses offered by colleges and universities. Consequently, we encourage the application of the “Seven Principles of Good Practice in Undergraduate Education” (American Association of Higher Education, 1987) to guide the design and implementation of Web-based courses. We offer concrete suggestions on how instructors can apply each principle to maximize the potential of Web-based technologies and promote positive learning outcomes in the virtual classroom.

During the last decade, advances in computer technology have increased individual access to the Internet. This increased access coupled with the development of educational course management software such as WebCT (http://www.webct.com), TopClass (http://www.wbtsystems.com), and Blackboard (http://www.blackboard.com) allows instructors to develop Web-based (and Web-enhanced) university-level courses. In 1997, almost 400 colleges augmented or supplanted conventional courses with online instruction (Velsmid, 1997). Furthermore, over 150 colleges and universities offered entire bachelor’s degree programs to students who rarely visit campus (Herther, 1997).

Despite the proliferation of Web courses, there is a paucity of empirical data on what constitutes effective pedagogy in the virtual classroom. For instance, the report “Distance Education: Review of the Literature” (Schlosser & Anderson, 1994) did not cite any Web-based course student outcomes research. A more recent review of distance learning research reported that only about 40 studies were empirical in nature and only a few involved courses taught via the Web (Merisotsis & Phipps, 1999).

The lack of empirical data on Web-based pedagogy is unfortunate because without an understanding of Web-based learning and instruction, there is the danger that Web-based course design becomes driven by technology rather than pedagogy (Trapp, Hammond, & Bray, 1996). In recognizing this gap between the research and the proliferation of distance learning courses, Chickering and Ehrmann (1996) recommended that distance-learning instructors should adopt the “Seven Principles of Good Practice in Undergraduate Education” developed by the American Association of Higher Education (AAHE; 1987). Although developed primarily for conventional classroom instruction, Chickering and Ehrmann argued that these principles are effective guides for the implementation of Web-based courses.

In this article we explore these principles with respect to their applicability to Web-based courses. Whenever possible, our analysis of the available research on Web-based courses is guided by these principles. We also discuss our own experiences with Web-based courses. Our combined experience includes the design and implementation of three different Web-based courses (Social Psychology, Statistical Methods in Psychology, and Research Methods in Psychology) across a total of 34 different sections. Finally, we offer suggestions on how instructors can implement each of these principles in the virtual classroom.

Principle 1: Encourage Contact Between Students and Faculty

In general, there are two forms of computer-mediated communication (CMC) in Web-based courses. The first, asynchronous communication, is time and place independent because the instructor and student can send messages to one another from different locations and times. E-mail and electronic fora (e.g., bulletin boards) are popular forms of asynchronous communication. In our experience, e-mail interactions quickly clarify information for students. For our courses, we typically respond within 1 hr to our students. Forum postings can also be an effective means for all members of the course to communicate with each other. Most forum postings from students in our courses tend to deal with issues of interest to all students (e.g., “Did anyone else have difficulty with problem #5?”). For this type of forum posting, instructors can provide clarification and detail to the entire class in a single posting (e.g., “You should try the formula on page 325.”). Instructor-initiated forum postings are also effective for addressing all students. Consequently, changes in deadlines, corrections to assignments, and alterations due to emergency situations can be dealt with quickly without waiting for the next class session. For example, during a recent hurricane evacuation we were able to post forum messages regarding canceled lectures and postponed homework assignments.

In general, we have found that we are able to reduce confusion and increase learning for students using asynchronous forms of communication. Furthermore, Wang and Newlin (2000) analyzed asynchronous communication and found that final grades in a Web class were correlated with the number of course forum postings read and written by students during the semester.
The second type of CMC, *synchronous communication*, is placed, but not time independent. This form of communication occurs in real time via electronic “chatrooms” (i.e., virtual meeting places) whereby the instructor and student agree to meet online at a specified time. This type of interaction affords additional lecture and discussion in Web-based courses. Although this lecturing format may appear problematic, we find it is close to a traditional lecture in a classroom. Instead of speaking and writing on the chalkboard, instructors type the lecture materials for the class to see immediately. Figures and diagrams that cannot be displayed in the chatroom can be available to students by posting them on a Web page prior to the lecture. Students can then print them out and refer to them during the lecture. Additionally, textbook figures can be effectively discussed when the instructor provides the necessary page numbers. Beyond reproducing the traditional lecture hall, the chatroom produces a marked increase in students’ interaction. In our experience, students seem much more willing to interact and ask questions during the chatroom lecture, possibly due to the increased level of anonymity. This heightened level of anonymity occurs even when students are identified by their names in chatrooms. Because students interact more, errors or unclear points are more likely to be corrected, and students are more likely to engage in a detailed review of the information. The interactivity of students in the electronic chatroom seems to allow for a greater amount of humor compared to conventional lectures. Humor seems to relax the students and make their interactions with the instructor more enjoyable. Indeed, we found that compared to other components of Web-based instruction, students rated chatrooms extremely high when responding to the statement “I found this component promoted my learning of the course material.” In a similar vein, Gunawardena and Zittle (1997) found that increasing the social presence of instructors via computer-mediated video conferencing led to greater student satisfaction and performance in class.

Wang, Newlin, and Tucker (2001) also showed that the quantity and type of synchronous communication can be predictive of Web-course outcomes. Specifically, a discourse analysis of course chatroom discussions revealed that final grades in the course correlated with the frequency by which a student responded to an instructor’s query and the number of times that a student was first to respond to such a query. Here, too, the technology seems to support the seven principles of undergraduate education, the understanding of course content, and student success in the class.

In general, Web instructors should use as many forms of CMC as feasible to encourage student–faculty contact. As Holmberg (1989) noted in his theory of guided didactic conversation, the core of teaching consists of the interaction between instructors and their students. However, this interaction should not merely be the accurate transmission of information from one party to the other. Just as importantly, the tone of the instructor’s communication should be personal and supportive, thereby encouraging the motivational and emotional involvement of the student with the course content (Holmberg, 1989).

Establishing personal contact with cyberstudents via CMC is especially important in light of recent research suggesting that individuals who spend extended amounts of time on the Web may become less socially involved and more depressed than prior to their involvement with the Web (Kraut et al., 1998). In this regard, the “loneliness” of the distance learner not only describes the experience of some cyberstudents, but may also relate to the experience of their Web instructors (Laird, 1999). Thus, encouraging a high degree of instructor–student interactivity is vital for the motivational and emotional involvement of Web instructors as well as their students.

Increased instructor–student interactivity is not without cost. The time and effort needed to maintain personal contact with students is considerably more in the virtual classroom compared to conventional class. In this regard we offer two suggestions. First, Web-based instructors should design course home pages with easily navigable layouts and unambiguous content. In this manner, instructors can reduce the e-mail traffic that occurs when a virtual classroom of students is confused about an item on the course home page. Second, if an error on the course home page is detected, instructors should immediately post a forum message indicating the nature of the error and its correction. Alternatively, if a class distribution list is available, the instructor can notify all of the students in the course with a single e-mail message.

**Principle 2: Develop Reciprocity and Cooperation Among Students**

Web-based CMC facilitates student-to-student interactions, thereby increasing student satisfaction and improving course outcomes. For example, Althaus (1997) and Meyers (1997) have shown the benefits of CMC for large conventional classes organized into small online discussion groups. The advantages of CMC are also evident for students in a Web-enhanced statistics laboratory (Varnhagen, Drake, & Finley, 1997). Although the authors did not provide outcome data in this study, it is notable that students rated the communication resources as more useful than the information resources when evaluating the Web components of their statistics laboratory.

As noted previously, without maintaining a high degree of student interactivity, Web instructors are risking the possibility that there will be a lack of motivational and emotional involvement on the part of their cyberstudents. One means of increasing student involvement is to use CMC as a powerful resource that can facilitate the development of cyberlearning communities. We have argued elsewhere (Wang & Newlin, 2000) that student collaboration may be more important for the virtual classroom compared to conventional classes. In fact, data reveal that students who regularly communicate with others (e.g., via e-mail, student-organized chatrooms, face-to-face contact) not only enjoyed these contacts, but had higher final grades in the class than distance learners who remained isolated (Hiltz, 1993; Kember, Lai, Murphy, & Yuen, 1992; Wang & Newlin, 2000).

Just as in the conventional class, instructors can use group projects to facilitate cyberstudent collaboration. However, to facilitate the formation of cyberlearning communities, Web instructors must be proactive in devising a Web-assisted means by which their students can contact one another to
Principle 3: Use Active Learning Techniques

Instructors can use the interactive and dynamic potential of the Web to promote active and critical thinking and to avoid student passivity and apathy (McKeachie, 1999). As noted previously, the instructor can and should use CMC to assist the formation of cyberlearning communities. When these learning communities are functioning well, study sessions are student moderated with some students serving as peer mentors and others as peer mentees. Goodlad (1998) and Twomey (1991) concluded that conventional peer mentoring benefits both mentors and their mentees. There are other ways to tap the potential of Web-based technologies. For instance, both high school and college students reported feeling very positive about their experiences when collaborating on the design of each other’s Web pages (Pychyl, Clarke, & Abarbanel, 1999). Anecdotally, we find that giving cyberstudents tips about Web-based search engines empowered them to do Web-based research even when a document search was not required for a class assignment. When we taught a Web-based course on Social Psychology, we found that many students would initiate their own Web research on the topic scheduled for discussion in a chatroom. For instance, prior to discussing the topics of prejudice and discrimination, several students downloaded the material that they found from “hate-group” Web sites. Subsequently, they were able to analyze this material in terms of psychological concepts such as the fundamental attribution error and in-group versus out-group bias.

In general, instructors can effectively use the vast amounts of available information on the Web to increase the active learning of cyberstudents. Instructors should take care to develop assignments and projects designed to take advantage of this existing potential. In our experience, students tend to become highly involved in the learning process when they seek out and identify relevant Web sites that are specific components of well-designed and organized assignments. Here, too, the technology provides a new mechanism to support important principles of undergraduate education.

Principle 4: Give Prompt Feedback

Web-based technologies allow for the electronic transmission of feedback as quickly as instructors deem desirable. For example, the instructor can send comments and grades on student assignments via e-mail or by forum postings (using coded student identities). In our experience, instructors can provide personalized, question-by-question feedback to each student in a timely manner. Furthermore, students are able to quickly communicate any questions that they might have about the feedback or request from the instructor greater explanation of any comments. Anecdotally, we find that this type of feedback seems to contribute to greater feelings of instructor contact, support, and availability for the students. Moreover, the response to students can be effectively augmented through the use of synchronous (e.g., cyberoffice hours) and asynchronous communication (e.g., group-specific e-mails, forum postings).

Instructors who are familiar with Web-based technologies can also be much more creative in their use of feedback. For instance, one common function of many Web servers is their ability to automatically monitor the frequency and type of contact made by remote users such as the students who enroll in Web-based classes (Shute & Regian, 1993; Svnum, Chen, & Bublitz, 1997). Web-based course management systems can therefore track students in terms of the number of times (and duration) that they access the various components of the virtual classroom (e.g., course home page, assignment links, course forum). Instructors who monitor this activity have additional information concerning the Web-based involvement of their students. This information is particularly valuable during the beginning of the semester as an early warning predictor of cyberstudent success (or failure) later in the course. For instance, Wang and Newlin (2000) showed that there is a strong correlation between the frequency of course home page hits during the first week of class and students' final grade in the class. Consequently, Web instructors should not hesitate to personally contact students who initially exhibit low involvement with the Web components of their class.

Principle 5: Emphasize Time-On-Task

The remoteness or distance of cyberinstructors may diminish any sense of urgency on the part of their students to spend time completing Web-course assignments. Although the research suggests that successful distance learners have an internal locus of control and are effective time managers (Atman, 1988; Dille & Mezack, 1991; Wang & Newlin, 2000), cyberinstructors cannot assume that these characteristics apply to all students. Consequently, techniques are needed to maintain student involvement throughout the semester. We have already discussed two such techniques; a high level of instructor–student contact and peer collaboration (e.g., cyberlearning communities) that can increase course performance and student satisfaction. To these techniques we offer two other suggestions: establishing regularly scheduled chatrooms (e.g., weekly discussions or lectures) and regular postings of graded assignments throughout the semester. Regularly scheduled chatrooms not only promote student (and instructor) involvement in the course, but also serve to increase the actual and perceived presence of the instructor in the course. By taking on an active role in chatrooms, the instructor can directly address, in a timely manner, issues that hinder students' time-on-task (e.g., content-based and technological problems). Moreover, to the extent that instructors assume the role of in loco parentis in
the virtual classroom, reminders about course assignments and deadlines can be relatively nonthreatening (as opposed to a personal e-mail reminder to a student).

Given the broad research support for distributed versus massed practice (Underwood, 1961), it is also crucial that cyberinstructors release (i.e., post on the Web) course materials on a regularly scheduled basis throughout the semester. These postings and deadlines should be specified on the course syllabus at the beginning of the term. Regular postings of course materials (and assignment deadlines) will prevent end-of-semester cramming, and just as importantly, give cyberlearning communities the opportunity to organize study groups as needed. Moreover, by not scheduling regular postings, the cyberinstructor has implicitly adopted the pedagogy that characterized the distance learning practice of an earlier technological era: the mail correspondence course.

Principle 6: Communicate High Expectations

In our experience, there are some students who misconstrue their first Web course as a correspondence course. This misinterpretation includes notions that course assignments are student paced, that there is minimal interaction between instructor and student, that Web-based courses are easy to master, and that peer collaboration is prohibited. However, as we previously noted, cyberinstructors should design their Web courses to facilitate a high degree of interactivity as well as to encourage students’ time-on-task with regularly scheduled postings and assignment deadlines.

Appropriate student expectations can be fostered only when instructors clearly communicate their own expectations for the course. How are student expectations fostered in the virtual classroom? Cyberinstructors can use at least four techniques to establish high expectations for their students. First, the course goals and objectives should be highly visible on the course home page. In this manner, enrolled students as well as prospective students can gain a perspective on the cyberinstructor’s expectations for the course. Second, the means by which students attain these goals and objectives, and the manner in which performance feedback is given, should also be clearly described on the course Web site (e.g., a grades link). Third, initial contact by the cyberinstructor (e.g., e-mail, chatroom, forum postings) should reiterate the goals and objectives specified on the course Web site. Fourth, feedback from the instructor throughout the semester should communicate performance expectations in a nonthreatening and supportive manner. Through this frequent, individualized feedback, any misconceptions about the required level of performance can be addressed and dispelled.

Although it is the responsibility of instructors to establish course expectations that are high, a question that many first-time cyberinstructors have is “How high is high?” Implicit in this question is the notion that the nature of distance-learning and Web-based technologies should somehow change the goals and objectives of the course. However, educational researchers have questioned whether there is any fundamental difference in the course outcomes of conventional versus distance-learning courses (Keegan, 1986; Russell, 1999; Shale, 1988). In our view, cyberinstructors should maintain the same expectations and standards for Web courses as they would for their conventional sections of the class. In fact, maintaining an equivalency of learning expectations across Web-based and conventional sections should be a guiding principle in the design and implementation of distance educators (Simonson, Schlosser, & Hanson, 1999). Adherence to this principle will ensure that Web-course design and implementation is driven by pedagogy rather than technology.

In practice, cyberinstructors can begin to adhere to this equivalency principle by employing a syllabus comparable to that used in a conventional class. Therefore, materials typically used by the instructor for a conventional class (e.g., textbook, assignments, lecture notes) would be similarly described on the Web-based syllabus. Furthermore, instructors should use lecture chatrooms when it is necessary to convey course content to augment reading assignments and asynchronous communication. As an additional means of quality control, cyberinstructors should also ensure that the graded components (e.g., quizzes and exams) of their Web course are similar to those given in conventional sections of the class.

However, cyberinstructors should not merely copy their class materials from a conventional hardcopy format into an electronic, Web-based format. To simply “shovel” course materials from one medium into another without attention to the potential (and pitfalls) of the second medium invites the inevitable and unflattering comparison with “shovelware” (Fraser, 1999). Instead of simply copying course materials from one medium to another, cyberinstructors should consider how they can attain course objectives that effectively take advantage of the technologies inherent in the Web.

For instance, Web course components that incorporate hypertext links are a fundamental means of tapping the potential of the Web (Beasley & Kent, 1996; Plaud, 1996). With judicious use of hypertext links (e.g., key concepts that are “hot linked” to a glossary), students can navigate through several course Web pages and concepts that might otherwise appear bewildering in their organization and relation to one another. In conjunction with graphical images, hypertext markup language (html) can be used to create “clickable” image maps whereby students can point and click on a location on an image to receive detailed information about that location (e.g., clicking on the image location of Broca’s area of the human brain to learn more about that area of the cortex).

Finally, as a means of quality control, cyberinstructors should maintain the principle of equivalency when devising the graded components of their Web class. Specifically, the quizzes, exams, and projects assigned in a Web course should be as similar as possible to conventional sections of the class. Consequently, cyberinstructors can compare course outcomes to assess the attainment of course goals and objectives.

Principle 7: Respect Diverse Talents and Ways of Learning

From the foregoing discussion, it should be evident that the creative use of Web-based technologies can support the diverse ways of knowing exhibited by cyberstudents. In addition to a textbook, cyberstudents can select from a rich array
of Web resources to master the content of a course (e.g., internal and external links, posted lecture notes, chatroom discussions, forum postings, student questions and comments). Well-designed pages on the course Web site (particularly the home page) incorporating a balance of hypertext links and graphical images will allow students who prefer either visualization or verbalization (Childers, Houston, & Heckler, 1985) to navigate with equal facility across the pages and content of the course.

With regard to demographic diversity, it should be noted that Web technologies can serve a democratizing function. Students with personal circumstances (e.g., physical disabilities, family responsibilities, distance, travel requirements) that prevent them from commuting to campus can take a Web course from their homes. Other students who would never consider speaking in a conventional classroom discussion may no longer be intimidated in an online chatroom. As we have already noted, students tend to be far more interactive in Web-based compared to conventional lectures. We have also found a significant reduction of noninteractive students in Web-based lectures. Finally, the merit of one’s contribution to a chatroom discussion is evaluated solely on the quality of one’s argument; there are no visible characteristics such as race, sex, disability, or ethnicity that may bias one’s attributions of the individual who has made a contribution to the chatroom discussion.

Summary

In this article we have demonstrated how the “Seven Principles of Good Practice in Undergraduate Education” (AAHE, 1987) can be applied to the design and implementation of Web courses. In applying these principles, we believe the development of Web courses is guided by sound pedagogical practice rather than driven by technology. Implicit in our application of these principles is the notion that educators should maintain equivalency in course outcomes regardless of whether instruction occurs in the conventional manner or in the virtual classroom.

In summarizing, we would like to emphasize a few issues that cut across many of the principles described previously. First, despite the heavy reliance on technology, personal contact of a highly interactive nature is still needed in the technologically rich environment of the Web. Rather than viewing the CMC as a remote and impersonal form of interaction, cyberinstructors should conceptualize CMC as a means of encouraging contact between themselves and their students. CMC is also a resource for facilitating the formation of cyberlearning communities. To ignore the technologies of CMC is to promote the loneliness of the distance learner and instructor.

Second, the successful design and implementation of Web courses can be a transformative experience in the way that students, as well as their instructors, approach the educational enterprise. Although we believe that course outcomes should be equivalent between conventional and Web-based courses, the experience of teaching and learning may be somewhat different (Simonson et al., 1999). In the virtual classroom, the lecturer’s lack of a podium and microphone promotes the democratization of class discussions, with the consequence that online chatrooms become a meeting place for lively and open discussion. Moreover, the format of chatroom discussions demands that instructors communicate in a concise and unambiguous manner. Cyberinstructors can not easily afford the luxury of “hand-waving” or digressing on a lecture topic in a chatroom. In fact, our experience has led us to believe that preparation of chatroom lectures and discussions has transformed, in a positive way, our approach to conventional classroom discussion. The increased interactivity in Web-based lecturing seems to ensure that students will inquire about awkward or confusing lecture points. The result is an increased level of clarity and understanding for the students.

Finally, we would like to reiterate that developing a Web course is not simply a matter of copying course material from one medium to another. If educators develop Web instruction solely as a means of changing student access, they have missed the point about using the Web as an instructional tool. Stated succinctly, “the extent to which a student gains the same pedagogical benefit from a printout of your Web resources as from the resources themselves is the extent to which you have done nothing of pedagogical value by using the Web” (Fraser, 1999, p. B8). Instead, cyberinstructors should think in terms of how Web-based technologies and CMC can promote desirable educational goals and outcomes (McKeachie, 1999).

References


Note
Send correspondence to Michael H. Newlin, Department of Psychology, University of Central Florida, 1519 Clearlake Road, Cocoa, FL 32922; e-mail: mnewlin@pegasus.cc.ucf.edu.
To test the effectiveness of an online introductory psychology course, we randomly assigned students to a large, traditional course or to an online course from a population of students who indicated that either course type was acceptable using a "waiting list" experimental design. Students in the online course performed better on exams and equally well on paper assignments compared to students in the traditional course. Online students also showed greater satisfaction with the course than those in the traditional course. Our results indicate that students who are amenable to taking either an online or a traditional course performed as well in an online course as students enrolled in a large, traditional course.

Institutions around the world have begun to offer online distance learning courses, and the number of students enrolled in online courses is projected to increase significantly. By the year 2006, 5 million students are projected to be enrolled in online courses (Symonds, 2001). The increase in the number of online course offerings is in part due to online Internet access. Students who cannot be physically present at a college campus for any reason, such as geographic constraints or disability, can enroll in online courses. In addition, online courses may allow more interaction between students and instructors than traditional courses. Although face-to-face contact may be limited or nonexistent, e-mail, chat, and threaded discussions can allow students and instructors to have more discussions than in traditional courses. Furthermore, online courses allow shy students to communicate without talking in front of others.

Although there are potential advantages of online course delivery, there are also potential disadvantages. For one, students and instructors may miss the face-to-face contact that takes place in traditional courses (Graham, 2001). Less face-to-face interaction may lead to feelings of isolation and higher attrition rates (Carr, 2000; Moore & Kearsley, 1996). Another potential disadvantage is that online courses rely heavily on technology. If a computer or network connection fails, the course cannot be reached.

Although there are potential advantages and disadvantages of online course delivery, it is not clear whether online courses are as effective as traditional courses. A report reviewing studies published in the 1990s found that distance learning students have similar grades and attitudes when compared with traditional students and that both students and instructors have positive attitudes toward distance learning (Institute for Higher Education Policy, 1999). Recent results from psychology courses seem to support this conclusion (Graham, 2001; Lawson, 2000; Waschull, 2001).

However, there are limitations of past research. In many studies, participants were not randomly assigned to different course modalities but had selected the type of course for themselves (Graham, 2001; Wegner, Holloway, & Garton, 1999). This lack of control is a major drawback to comparing online and traditional classes because students selecting online courses may be different than students choosing traditional courses. In addition, the method for assessing learning is often not comparable between online and traditional courses. Typically, students in traditional courses take public, proctored exams, whereas students in online courses often take unproctored exams privately (e.g., Waschull, 2001), potentially giving online students an advantage due to the accessibility of books, notes, and even other people. To compare learning outcomes between online and traditional courses appropriately, students should take identical exams under the same conditions.

We designed our study to address these methodological issues. First, we created a waiting list made up of students who were admitted to an introductory psychology course. From this waiting list, we randomly assigned students to either the online or traditional section. Second, we required the online and traditional students to take identical, proctored exams at the same time in a traditional classroom. We compared the effectiveness of the two courses on grades and student evaluation ratings.

**Method**

**Participants**

Twenty-three undergraduate students enrolled at a large state university participated. From the pool of volunteers, the instructor randomly assigned 11 students to the large (N = 477), traditional course and 12 students to the online course. Two of the 11 students assigned to the traditional course dropped the course within the first 2 weeks of class and were not used in the analyses.
Traditional Course

The traditional class met for 75 min twice per week. The meetings consisted of a combination of lecture, video, discussion, activities, in-class writing assignments, and exams. In addition, the instructor required students to read a textbook, participate in three online discussions, and write three short papers. Students used WebCT software to post two comments per online discussion.

Online Course

Students met in an online classroom several times per week for 15 weeks. Students accessed the course via the eCollege course delivery platform, which consists of content, communication, and assessment tools. The content tools allowed students to view the course material and assignments. The students had no traditional lectures. However, unit introductions consisting of text and audio files, learning objectives, reading assignments, Web activities, short paper assignments, and practice exams were posted on the eCollege site. Students used the same textbook as in the traditional class. Students interacted with other students and the instructor via e-mail and threaded discussion. Students and the instructor participated in two online discussions each week. The instructor also gave the students general feedback regarding the course content at the end of each week.

In a typical week, the instructor required online students to read a chapter, participate in two online discussions, and complete several Web activities. Furthermore, during the course of the semester, they wrote the same paper assignments as the traditional students. The only time online students met their classmates and the instructor face to face was to take the four exams with students in the traditional course.

Measures

Participants completed four exams, three two-page papers, and a course evaluation. The exams consisted of 28 multiple-choice questions. We used 23 of the 28 questions, identical between courses, in the analyses. Teaching assistants graded the papers on a scale from 1 (poor) to 7 (excellent). To test the students’ evaluations of the two courses, the experimenters constructed four scales. The four scales consisted of instructor evaluation, overall course rating, interaction and feedback rating, and amount of time spent on the course (see Table 1). The course evaluations consisted of approximately 30 items, most of which were rated on a scale ranging from 1 (negative) to 7 (positive).

Results

Exams

A mixed-design ANOVA of students’ exam grades, with course (online vs. traditional) as a between-subjects factor and exam (first, second, third, or fourth) as a within-subjects factor, revealed a significant main effect for course format, $F(1, 19) = 4.96$, $p < .05$, Cohen’s $d = 0.57$. Overall, the online students ($M = 79.63$, $SD = 12.95$) correctly answered a higher percentage of exam questions than the traditional students ($M = 71.93$, $SD = 14.11$). Neither the main effect for exam nor the interaction was significant.

Papers

A mixed-design ANOVA on students’ paper grades, with course (online vs. traditional) as a between-subjects factor and paper (first, second, or third) as a within-subjects factor, revealed no significant main effect for course format, $F(1, 19) = 0.09$, $p > .05$. The online ($M = 6.85$, $SD = 0.33$) and tra-
ditional ($M = 6.86, SD = 0.25$) students had nearly identical mean paper grades. Neither the main effect for paper nor the interaction was significant.

**Course Evaluations**

A mixed-design ANOVA, with scale as the within-subjects variable and course (online vs. traditional) as the between-subjects variable, revealed a significant main effect for course format, $F(1, 19) = 7.39, p < .05$. The course evaluations were positive for both courses, but the online students gave more positive ratings than the traditional students. To test the difference between the two courses further, we analyzed each scale separately with independent group $t$ tests.

The online students ($M = 6.56, SD = 0.47$) gave significantly higher ratings of the instructor than the traditional students ($M = 5.78, SD = 1.09$), $t(19) = 2.24, p < .05$, Cohen’s $d = .93$, and the online students ($M = 6.24, SD = 0.62$) rated the amount and quality of interaction and feedback higher than the traditional students ($M = 4.76, SD = 1.45$), $t(19) = 3.20, p < .01$, Cohen’s $d = 1.33$. For overall course rating, there was no significant difference between the online students ($M = 6.15, SD = 0.51$) and the traditional students ($M = 5.22, SD = 1.61$), $t(19) = 1.88, p > .05$, Cohen’s $d = .77$.

The students reported spending approximately the same amount of time on the courses, which included listening to lectures, reading, studying, writing papers, completing assignments, and participating in discussions, $t(19) = 1.56, p > .05$, Cohen’s $d = .65$. On average, the online and traditional students reported spending 5 hr on their course each week. Students in both courses reported spending about 2 hr per week reading the textbook and approximately 4 hr studying for each exam.

**Discussion**

The results suggest that students enrolled in online courses perform at least as well as students enrolled in traditional courses. Our online students performed better on exams than the traditional students, even though students from both courses reported spending about the same amount of time studying. Both groups of students earned high grades on the paper assignments. The majority of students gave high course evaluation ratings, but the online students reported being more satisfied with the instructor and with the amount of interaction and feedback they received.

Although previous researchers found similar grades and course evaluation ratings between traditional and online courses (Graham, 2001; Institute for Higher Education Policy, 1999; Lawson, 2000; Wegner et al., 1999), their findings are problematic because students were not randomly assigned to the different courses, meaning that students who self-selected online courses may have been more motivated, more intelligent, or more familiar with the online environment. Second, in other research (e.g., Waschull, 2001), the online students did not take proctored exams, which may have given them a test-taking advantage.

Our research employed a procedure to help ensure comparability between the online and traditional groups. The results showed that online students performed at least as well as traditional students and in some ways better. Furthermore, the online students reported that they were generally pleased with their online learning experience and stated that they would take another online course if given the opportunity.

However, there are limitations to the current research. First, the pool of participants was made up of those students wishing to enroll in introductory psychology but who voluntarily agreed that either course modality was acceptable to them. It is possible that students who wished to enroll in introductory psychology, but who did not agree in advance that either option was acceptable, might yield different results. In an effort to address this issue, grades of students in the large lecture class who had agreed to participate in either modality were compared to grades in a sample of those who had not indicated willingness to participate in either modality. There was no significant difference between the two samples, suggesting that the participants in the study were not atypical of the class as a whole.

Second, the comparison between the online and traditional course modality may be affected by the size difference between the two classes. Students in the traditional class were members of a large, lecture course, whereas the online course was small. It is possible that the greater student satisfaction in the online course may be related to size rather than the course delivery method. In addition, the relatively small numbers of students in the study limited its power to discern statistical differences; any null findings may be a reflection of the low numbers.

Still, the results of this study clearly add support to the body of research that show that online distance learning courses are at least as effective as traditional courses. The results suggest that students who are amenable to taking courses online can perform as well as students enrolled in large, traditional lecture courses.

Ultimately, though, the success of students in a particular class may be due less to the instructional medium than to the nature of activities (e.g., online discussions, class participation) carried out by students and instructors in the course (Bartlett & Strough, in press). As Ragan (1998) argued, “good teaching is good teaching” (p. 1), and the outcome for students of psychology may be more a function of what they do, rather than how they do it.

**References**


Notes

1. We are grateful to Kevin Aiken and the University of Massachusetts–Amherst Division of Continuing Education for support of this research.

2. We also thank David Rondina for helping in the collection of data.

3. Send correspondence to Robert S. Feldman, Department of Psychology, University of Massachusetts, Amherst, MA 01003; e-mail: feldman@psych.umass.edu.


Notes

1. Pam Marek is now at Kennesaw State University.
2. Portions of this article were presented at the 15th Annual South-eastern Conference on the Teaching of Psychology, Kennesaw, GA, February 2003.
3. We thank Randolph Smith and three anonymous reviewers for their comments on an earlier version of this manuscript.
4. Send correspondence and requests for copies of current assignments or grading checklists to Pam Marek, Department of Psychology, #2402, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144–5591; e-mail:(pmarek@kennesaw.edu.

Enhancing Online Instruction With Humor

Frank M. LoSchiavo and Mark A. Shatz
Ohio University—Zanesville

We tested the effectiveness of humor as an instructional strategy in an online general psychology course. We randomly assigned students to a standard section of the online course or to a section enhanced with additional humorous content. Humor significantly influenced student interest and participation but had no effect on overall course performance. We discuss the implications of these results, and we provide suggestions for integrating humor into online course components.

Both teachers and students believe that learning should be fun. Teachers rely on humor to create warm, inviting classroom environments (e.g., Buskist, 2002; Roth, 1997), and students report that humor increases how much they like their courses, their instructors, and how much they learn (e.g., Gorham, 1988; Wanzer & Frymier, 1999; White, 2001). Even though objective measures of learning (e.g., exam scores) have produced mixed results, most researchers have concluded that humor is an effective instructional strategy (e.g., Berk, 2000; Berk & Nanda, 1998; McMorris, Boothroyd, & Pietrangelo, 1997; Oppliger, 2003; Zillman & Bryant, 1989; Ziv, 1988).

However, humor is not a pedagogical panacea; no research has evaluated whether humor is beneficial in online teaching environments, where most communication is text-based and there is little or no real-time interaction between instructors and students. Using an experimental design, we tested the effectiveness of humor in an online general psychology course. We hoped that humor would help establish a comfortable rapport that would ultimately alter the immediacy (i.e., perceived social distance; Kelley & Gorham, 1988) of the instructor to the students. Previous research has shown that humor and immediacy are positively related to course satisfaction (Gorham, 1988) and class participation (Korobkin, 1988). Thus, we predicted that students exposed to additional humor would find the course more interesting and enjoyable, and we suspected they would use the course Web site more frequently. Because objective measures of learning have produced inconsistent results (McMorris et al., 1997), we did not expect to find differences in academic performance, particularly because the additional humor was not specifically intended to strengthen one’s conceptual understanding of the material.

Method

Participants

Participants were 44 undergraduates (38 women, 6 men) who registered for an online general psychology course.

Procedure

After registration closed, we randomly assigned students to one of two sections of the course, each delivered over the Internet using the Blackboard Learning System (Version 5.5). Student interaction between sections was unlikely because the online students rarely visited campus and a log-in procedure restricted access to one’s assigned section.

Both sections featured learning objectives, chapter summaries, text-based lectures, quizzes, and discussion boards. Although we did not intend for the standard section to be humorous, the lectures had a lighthearted, casual tone, making them somewhat entertaining to read. To create the humor-enhanced section, we systematically incorporated humor into the existing course components by adding two or three content-relevant jokes to each of the 26 lectures, humorous cartoons to each of the 14 quizzes, and witty remarks to the 10 electronic announcements. Thus, the only difference between the two sections was that one featured overtly humorous material.

Students followed a strict schedule that included weekly reading assignments, mandatory class participation (using discussion boards), and brief quizzes. In addition, students completed four proctored exams, including a comprehensive final exam. To assess academic and social behavior, we tracked how often students accessed course contents (e.g., lectures) and features (e.g., e-mail). On completion of the quarter, we sent students a brief survey to evaluate how humor influenced their perceptions of the course.

Results

Based on prior research, we estimated that effect sizes would be relatively small ($d \approx 0.30$; Cohen, 1988). Consequently, we chose to compensate for low statistical power by performing one-tailed tests.
Academic Performance

We based final grades on class participation, quiz scores, and exam scores. The results appear in Table 1. As predicted, there were no differences in final grades between the two sections. However, students assigned to the humor-enhanced section earned significantly more class participation points, \( t(42) = 2.19, p < .05 \). This analysis demonstrates that students in the humor-enhanced section were more likely to participate in online discussions.

Course Usage

To better understand trends in course usage, we sorted each segment of the course into one of three categories (lecture, social and interactive, or organizational) and analyzed corresponding hit rates. The lecture category included sections for lectures, topic previews, and learning objectives. The social and interactive category included sections for email and discussion boards. The organizational category included sections for course policies, assignments, and current grades. The results appear in Table 2.

Although there were no significant differences in overall course usage, students assigned to the humor-enhanced section used social and interactive features more frequently, \( t(42) = 1.71, p < .05 \). Informal observations also supported this finding. For example, students in the humor-enhanced section were more likely to post personal messages and to reply to other students’ questions.

Students’ Perceptions of the Course

Sixty-four percent of the students returned the anonymous survey. The results appear in Table 3. Students in the humor-enhanced section were more likely to appreciate the humorous content, recommend that humor continue to be used, and agree that humorous content made the course more interesting.

Subsequent Course Selection

We predicted that additional humor would spark added interest in online courses and in psychology courses. However, an analysis of registration records yielded no significant differences between the two groups. The results appear in Table 4.

Discussion

These findings provide direct support for our general hypothesis that pedagogical humor enhances online instruction. Although online learning environments lack many of the interpersonal dynamics of traditional instruction (e.g., face-to-face communication), this study demonstrates that humor can help create an electronic atmosphere that fosters both academic interest and social relations. Specifically, students in the humor-enhanced section were more likely to agree that humorous content made the course more interesting, and they interacted with each other more often.

Nevertheless, additional humor did not increase overall performance. Although humor can foster a positive online environment, it is not a substitute for traditional scholarship. For example, when discussing dreams, an instructor might joke that Martha Stewart dreams about designing trendy prison jumpsuits. Although the joke is easy to understand, it lacks conceptual relevance, so it is unlikely to help students better understand the psychology of dreams (for a related discussion, see Mayer, 2001). In fact, we worry that humor may pacify confused students, leaving them less likely to seek clarification when necessary. Others have voiced similar concerns. For example, Olson and Clough (2003) noted that a potential consequence of the “education as entertainment” movement is that students may develop the belief that learning is easy. Although additional research is needed in this area, evidence suggests that conceptual humor fosters retention (Kaplan & Pascoe, 1977).

It is also possible that our humor manipulation was too weak to influence academic performance. Although the humor-enhanced section featured overt humor, the data indicated that students enjoyed the lighthearted, casual tone in the standard section as well. In the future, researchers may want to test if stronger manipulations influence academic performance.

Nevertheless, the benefits of pedagogical humor merit its use as an instructional strategy. When incorporating humor into online instruction, instructors should consider the following issues. First, the majority of the humor should address a specific instructional objective. After all, the primary goal is

Table 1. Student Performance by Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Standard</th>
<th>Humor-Enhanced</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation</td>
<td>92.21</td>
<td>12.45</td>
<td>98.38</td>
<td>4.36</td>
<td>4.36</td>
<td>2.19*</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Quizzes</td>
<td>59.86</td>
<td>7.36</td>
<td>60.50</td>
<td>6.65</td>
<td>0.30</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exams</td>
<td>72.37</td>
<td>12.27</td>
<td>72.10</td>
<td>10.63</td>
<td>0.09</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final course grade</td>
<td>76.76</td>
<td>11.13</td>
<td>76.71</td>
<td>9.69</td>
<td>0.02</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Means represent the average percentage of total points earned in each category. \( d = \) Cohen’s standardized effect size.

\( \ast n = 22 \)

\( \ast p < .05 \), one-tailed.
Table 3. Students’ Perceptions of Course-Related Humor

<table>
<thead>
<tr>
<th></th>
<th>Standarda</th>
<th>Humor-Enhanceda</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>I appreciated that humor was used in the course.</td>
<td>4.14</td>
<td>1.10</td>
</tr>
<tr>
<td>The use of humor made the course more interesting.</td>
<td>4.21</td>
<td>0.89</td>
</tr>
<tr>
<td>The use of humor made the course more enjoyable.</td>
<td>4.36</td>
<td>0.84</td>
</tr>
<tr>
<td>I was motivated to read the course material because of the humor.</td>
<td>3.71</td>
<td>1.14</td>
</tr>
<tr>
<td>I recommend that humor continue to be used in the course.</td>
<td>4.21</td>
<td>1.05</td>
</tr>
<tr>
<td>I recommend that more humorous material be added to the course.</td>
<td>4.00</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Note. Students responded to each item using a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). d = Cohen’s standardized effect size.

*p < .05, one-tailed.

Table 4. Credit Hours by Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Standarda</th>
<th>Humor-Enhanceda</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Online credits</td>
<td>1.00</td>
<td>2.62</td>
</tr>
<tr>
<td>Psychology credits</td>
<td>2.45</td>
<td>3.43</td>
</tr>
<tr>
<td>Total credits</td>
<td>35.55</td>
<td>21.47</td>
</tr>
</tbody>
</table>

Note. Means represent the average number of quarter hours in each category during the academic year following enrollment in the present course. No differences between the two groups were statistically significant. d = Cohen’s standardized effect size.

References


Note

Send correspondence to Frank M. LoSchiavo, Department of Psychology, Ohio University–Zanesville, 1425 Newark Road, Zanesville, OH 43701; e-mail: loschiav@ohiou.edu.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder’s express written permission. However, users may print, download, or email articles for individual use.
COMPUTERS IN TEACHING

Predicting Success in Online Psychology Courses: Self-Discipline and Motivation

Stefanie B. Waschull
Athens Technical College

This article addresses factors associated with student success in online psychology courses. Prior to beginning an online course, students completed measures of self-discipline and motivation, time commitment, study skills, preference for text-based learning, access to technology, and technology experience. Schrum and Hong (2002) proposed that these student characteristics predict online course success. I used scores on these factors to predict student performance in online introductory psychology and online human growth and development courses. Self-discipline and motivation was the only factor predictive of online psychology course success. My results contradict the model proposed by Schrum and Hong but are consistent with research on the role of motivation in success.

During the past decade, distance education programs have proliferated (Council for Higher Education Accreditation, 1999). Many traditional colleges and universities added distance education offerings to their programs. Furthermore, new, private, for-profit institutions began offering competitive distance education programs.

There is considerable research comparing the effectiveness of distance education and online courses to traditional courses (Beare, 1989; Fox, 1998; McKissack, 1997; Sonner, 1999; Waschull, 2001). The general consensus is that there are no significant differences in effectiveness between distance learning and traditional learning techniques. In an early study, Beare (1989) compared the effectiveness of videotape, audiotape, and telelectures to traditional courses and found no significant differences in student grades. More recently, McKissack (1997) showed that distance education students from a variety of colleges and universities did not differ from traditional students at those institutions on grade point average. Waschull (2001) compared students in online introductory psychology courses with similar students in traditional courses and found that the two groups did not differ in terms of test performance. Furthermore, Fox (1998) surveyed the research on distance education and found no empirical studies that indicated distance education students were deficient in skills compared to traditional students.

However, providing online distance education in an increasingly competitive environment raises important issues. There is a risk that institutions may focus on the mechanics of providing online courses to as many students as possible without adequate consideration of the pedagogical soundness of online materials or the reality of whether students are equipped to perform successfully in online courses. For example, Bonk and Dennen (1999) pointed out that popular Web-based courseware programs primarily provide administrative tools for faculty to post lectures notes and quizzes, provide Web links, and deliver assignments. Bonk and Dennen concluded these administrative tools do little to support student critical thinking or help students generate knowledge.

In an effort to address such concerns, Schrum and Hong (2002) identified organizational, pedagogical, institutional, and student factors they believed were related to the success of online courses. They provided students with a substantive needs assessment at more than 30 institutions that offered postsecondary online learning opportunities. Based on their analysis, Schrum and Hong identified seven critical factors believed to be related to student success in the online environment: personal traits such as self-discipline, lifestyle factors such as adequate free time to commit to the course, motivation to perform well in the course, strong study skills, a preference for text-based learning, reliable access to technology, and technology experience prior to the course. Schrum and Hong proposed that the seven student characteristics are valid predictors of student success based on an analysis of existing measures and the opinions of instructors. Ratings provided by experienced online instructors indicated that they agreed that access to technology and experience with technology were important for student success. Experienced instructors also agreed that adequate time commitment and preference for text-based learning were important, and most instructors agreed that motivation was an important factor in determining success. Generally, the experienced online instructors did not identify study skills and personal characteristics as important for student success. Despite the low ratings given to study skills and personal characteristics, instructors frequently commented that self-discipline was one of the most important factors for determining student success. Although the opinions of experienced online instructors give useful insight into the importance of these characteristics, I was interested in establishing the reliability of the factors identified by Schrum and Hong and determining whether they were predictive of student success in online psychology courses.
Method

Materials

I developed an online questionnaire designed to measure the seven student characteristics: personal traits, lifestyle factors, motivation, study skills, a preference for text-based learning, access to technology, and technology experience. I developed 23 items to measure these characteristics—3 or 4 items for each factor. I based several of the items on questions developed by Schrum (2003) to help students determine if they were ready to take an online course. Items addressed the degree to which students were able to meet the demands of an online course. For example, I asked students whether they could devote 10 to 20 hr a week to the course, whether they had the required computer technology. Items asked participants to rate the degree to which they agreed or disagreed that the item described them using a scale from 1 (I disagree completely) to 5 (I agree completely).

Participants and Procedure

Participants were either enrolled in one of two sections of online introductory psychology or in one section of online human growth and development. During the first week of the academic quarter students completed the questionnaire online. I sent a follow-up e-mail to those students who did not complete the questionnaire within 24 hr. Participation was voluntary. I asked 86 students to participate; 57 (66%) completed the questionnaire. Participants ranged in age from 18 to 46; 88% were White, 12% were African American, and 73% were women.

Results

Participation and Attrition

The frequency of withdrawals was not different for students who completed the questionnaire when compared to students enrolled in the three online sections, $\chi^2(1, N = 57) = .001, p > .05$. Furthermore, the distribution of race and sex was similar for the students enrolled in the online courses with those who completed the questionnaire: race, $\chi^2(1, N = 57) = .004, p > .05$; sex, $\chi^2(1, N = 57) = .17, p > .05$. The mean age of those who completed the questionnaire was not significantly different than the mean age of students enrolled in the online courses (M = 24.01, SD = 7.41) was not significantly different than the mean age of students enrolled in the online courses (M = 24.18, SD = 7.23), $t(57) = .02, p > .05$.

Reliability

I summed items reflecting each of the seven characteristics to form subscales and conducted a reliability analysis. I eliminated nine items from subscales and from the overall measure because the coefficient alpha was higher without the item. There were a total of six subscales: (a) self-discipline/motivation (SD/MOT)—“I usually meet all my deadlines,” “I am able to exercise self-discipline,” and “I am highly motivated to do well in my online course,” coefficient $\alpha = .49$; (b) adequate time commitment (TIME)—“I can devote 10 to 20 hr a week to my online course,” “The extra demands of my education place a strain on my relationships at home or at work,” coefficient $\alpha = .47$; (c) study skills (SS)—“I feel I can learn well on my own,” “Sometimes I think I know the material but then I don’t do well on the test,” coefficient $\alpha = .58$; (d) preference for text-based learning (TEXT PREF)—“I have good reading comprehension,” “I need to hear something repeated before I remember it well,” “I often need to ask an instructor to repeat directions or explain some aspect of an assignment,” coefficient $\alpha = .54$; (e) access to technology (TEC ACC)—“I have access to a computer that can access the Internet almost all the time in my home,” “I can access e-mail almost anytime in my home,” coefficient $\alpha = .91$; (f) technology experience (TEC EX)—“I can download new software programs for my computer,” “Learning new technology is not my idea of fun,” coefficient $\alpha = .44$. Note that I combined the subscales of self-discipline and motivation. The coefficient alpha for the motivation subscale was .24 and the coefficient alpha for self-discipline was .33. One item from the motivation subscale combined with the two items on the self-discipline subscale produced a coefficient alpha of .49. This combination was appropriate because self-discipline and motivation were the only two personality factors on the measure.

Online Course Performance

The four measures of online course performance were test score average, assignment average, final exam score, and final course average. Tests consisted of about 50 multiple-choice questions and two or three essays, assignments were two- to three-page papers about topics covered in the course, and the final exam was a cumulative test with 60 to 70 multiple-choice items. I did not include the lowest test score and the lowest assignment score in the final average. To determine the final course average, I weighted the test and assignment grades by .8 and final exam grade by .2.

Table 1 shows that, of the six subscales, only self-discipline/motivation was significantly correlated with test score average, $r(55) = .44, p < .001$; assignment score average, $r(55) = .29, p < .05$; final exam score, $r(55) = .36, p < .01$; and final course average, $r(55) = .43, p < .001$. Three of the measures were nonoverlapping measures of student performance (test score average, assignment average, final exam score), whereas final course average was an aggregate of the other measures and was included to serve as an overall indicator of students’ performance. Not surprisingly, participants’ scores on the measures were correlated.

Discussion

Based on the first-order correlations, six of the seven factors proposed by Schrum and Hong (2002) did not appear to be correlated with course performance. Only self-discipline/moti-
vation significantly correlated with test score average, assignment average, final exam score, and final course average. These results raise several important issues. First, because the efficacy of the measure for predicting success in online psychology courses was the subject of exploration, I retained subscales with low alphas in these exploratory analyses. My findings are preliminary until confirmed using measures with greater reliability. If the measures had greater reliability, factors other than self-discipline/motivation might also predict online course success. Second, it is interesting to note that factors such as access to technology and technology experience did not relate to online course performance but self-discipline did. These findings are not consistent with Schrum and Hong's (2002) claim that technology experience and access are important for online course success. On the other hand, these results are not surprising in light of the widely accepted contention that motivation plays an important role in student success (Dweck, 1986; Elliot & Dweck, 1988). These results raise questions about whether the factors predicting online course success are any different than factors that predict success in regular psychology courses. Indeed, the possibility that factors predicting success in online courses are not different than factors predicting success in traditional courses is consistent with the finding that student performance in distance education courses is not significantly different than in traditional courses (Beare, 1989; Fox, 1998; McKissack, 1997; Sonner, 1999; Waschull, 2001). I intend to address the comparable role of self-discipline/motivation in online and traditional courses in further research.

### References


---

#### Table 1. Intercorrelations Between Student Characteristics and Performance

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test average</td>
<td>—</td>
<td>.63*</td>
<td>.55**</td>
<td>.83**</td>
<td>.44**</td>
<td>.06</td>
<td>-.05</td>
<td>-.16</td>
<td>-.03</td>
<td>.10</td>
</tr>
<tr>
<td>2. Assignment average</td>
<td>—</td>
<td>—</td>
<td>.25</td>
<td>.55**</td>
<td>.29*</td>
<td>-.05</td>
<td>-.14</td>
<td>-.27</td>
<td>-.12</td>
<td>.14</td>
</tr>
<tr>
<td>3. Final exam</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.77**</td>
<td>.36**</td>
<td>-.08</td>
<td>.03</td>
<td>-.01</td>
<td>.01</td>
<td>.19</td>
</tr>
<tr>
<td>4. Final average</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.43**</td>
<td>.01</td>
<td>-.08</td>
<td>-.10</td>
<td>.04</td>
<td>.15</td>
</tr>
<tr>
<td>5. SD/MOT</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.06</td>
<td>.06</td>
<td>.20</td>
<td>.14</td>
<td>.36**</td>
</tr>
<tr>
<td>6. TEC EX</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-.27*</td>
<td>.16</td>
<td>.10</td>
<td>.03</td>
</tr>
<tr>
<td>7. TEC ACC</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.09</td>
<td>-.07</td>
<td>.21</td>
</tr>
<tr>
<td>8. TEXT PREF</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.32*</td>
<td>.24</td>
</tr>
<tr>
<td>9. SS</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.03</td>
</tr>
<tr>
<td>10. TIME</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

**Note.** SD/MOT = self-discipline/motivation; TEC EX = technology expertise; TEC ACC = access to technology; TEXT PREF = preference for text-based learning; SS = study skills; TIME = adequate time commitment.

*p < .05. **p < .01.

---

Send correspondence to Stefanie B. Waschull, Athens Technical College, 800 Highway 29 North, Athens, GA 30601; e-mail: swaschull@athenstech.edu.
Are Computer-Assisted Teaching Methods Effective?

Kurt A. DeBord  
Lincoln University

Mara S. Aruguete  
Stephens College

Jeannette Muhlig  
Missouri Department of Social Services

Jefferson City, Missouri

Two studies examined effects of computer-assisted (CA) teaching methods in introductory psychology classes. In Study 1, we provided students with lectures supplemented with either overhead transparencies or CA visuals. In Study 2, we compared students who used an optional Web site with students who did not. In both studies we held constant lecture content, course instructor, exams, and assignments. Results of the two studies showed that students liked the CA teaching interventions, although CA instruction had no effect on student performance in the courses. Based on these and other published findings, we recommend that universities examine closely their goals and priorities when devoting resources to instructional technology.

Teaching that incorporates computer tools has seen phenomenal growth over the past two decades. However, the development of classroom technology has outpaced the research published in assessing its effectiveness. Over a decade ago, Welsh and Null (1991) reviewed the research on computer-based instruction and concluded that empirical efforts were desperately needed to inform educators about the efficacy of computer-assisted (CA) teaching. Today, efficacy studies continue to be in short supply, despite calls for such scholarship (Hall, Watkins, & Ercal, 2000; Johnstone & Forsyth, 1996).

Publications on technology use in psychology classes emphasize the development of computer-based materials but devote less attention to how those materials affect student performance. CA teaching has been used creatively in a range of undergraduate psychology courses (Aberson, Berger, Healy, Kyle, & Romero, 2000; Brothen, 1997; Graham, 1997; Langston, 1998; Matthews, 1999; Neuhoff, 2000; Newcomb, Berkebile, Newman, & Parker, 1998; Plous, 2000; Riniolo, 1997; Sherman, 1998; Varnhagen, Drake, & Finley, 1997). Researchers typically find that student evaluations of CA instruction are overwhelmingly favorable (Aberson et al., 2000; Brothen, 1997; Graham, 1997; Newcomb et al., 1998; Plous, 2000; Sherman, 1998; Varnhagen et al., 1997). Although student evaluations are essential in judging the worth of a teaching tool, a number of researchers have gone a step further by investigating performance outcomes (e.g., grades). The results of these studies have been mixed, making it difficult to draw clear-cut conclusions about the efficacy of CA teaching (Erwin & Rieppi, 1999; Forsyth & Archer, 1997; Kazmerski & Blasko, 1999; Stoloff, 1995; Welsh & Null, 1991; Worthington, Welsh, Archer, Minde, & Forsyth, 1996).

Our university invested in sophisticated classroom computer and projection equipment with the goal of improving instruction. We conducted the following studies to investigate the efficacy of teaching with this equipment. The first study examined the use of computer presentation software versus overhead projector presentation methods. The second study investigated the potential benefits of providing a course Web site as a resource for students.

Study 1

Given the mixed results of previous research, the direction of our hypotheses was influenced by the hope that our university’s investment in teaching resources was well spent. For Study 1, our hypotheses were (a) students in the CA condition would perform better than students in the traditional condition on final exam scores and overall course grades when American College Test (ACT) scores were statistically controlled, and (b) students in the CA condition would prefer CA presentations more than any other presentation style whereas students in the traditional condition would not prefer CA presentations.

Method

Participants. Eighty-three students were enrolled in one of two sections of an introductory psychology course at a small (approximate enrollment = 3,500), Midwestern, historically Black university with an open-admissions policy. The sample included 52 women and 31 men. Fifty-one percent of the sample reported being White, non-Hispanic; 42% African American, non-Hispanic; 2% Hispanic; 1% Asian American; and 4% other. Most participants (84%) were first-year students. Participants’ ages ranged from 18 to 39 with a mean age of 20.33 (SD = 3.82).

Design and procedure. We compared two sections of the same course. One section (n = 42), the traditional condition, received a standard lecture format with overhead transparencies as the primary lecture supplement. The CA section (n = 41) received a standard lecture format with Microsoft PowerPoint slides as the primary lecture supplement. The content of the overhead and computer-generated slides was similar, although the latter incorporated animation and sound into the presentations. Course instructor, lecture content, exams, and assignments were identical for the two sections. Course grades represented the mean of 10 classroom exams and 11 homework assignments. Because university policy specifies
that course changes are not allowed after the first week of the semester, students were not allowed to change sections after the first week of class. The traditional section met 1 hr prior to the meeting of the CA section. Both sections met three times a week for a 16-week semester. All students completed questionnaire packets at the beginning and end of the semester. The packets contained a demographics questionnaire and a survey that asked students to indicate which of the following types of presentation formats they preferred: lecture only, lecture with overhead slides, lecture with computer slides, or other.

**Results**

As a check for comparability of the groups, we compared the two conditions on ACT scores. There was no significant difference between the CA group (M = 17.51, SD = 4.18) and the Traditional group (M = 18.51, SD = 2.99), t(74) = 1.54, p = .33, η² = .03, observed power = .33. Hypothesis 1 predicted that students in the CA condition would earn higher grades than students in the traditional condition. ANCOVA of final exam grades indicated that the CA group (M = 80.99, SD = 12.51) did not differ when ACT scores were statistically controlled, F(1, 73) = .001, p = .98, η² = .00, observed power = .05. Similarly, ANCOVA of final course grades showed that the CA group (M = 79.22, SD = 13.02) and the traditional group (M = 80.99, SD = 12.51) did not differ when ACT scores were statistically controlled, F(1, 73) = .001, p = .98, η² = .00, observed power = .05. In short, the two groups did not significantly differ in academic performance.

Hypothesis 2 predicted that at the end of the semester students in the CA section would prefer CA presentations, whereas the students in the traditional section would not indicate a preference for such presentations. None of the students in the traditional section preferred CA presentations, χ²(3, N = 36) = 54.89, p < .001. Among students in the CA section, nearly one third preferred the CA presentation format χ²(3, N = 37) = 8.08, p = .04. No student in either section preferred CA presentations at the beginning of the semester. Although these results appear to support Hypothesis 2, they must be qualified by the possibility that students in the traditional section may have had no exposure to CA presentations and, thus, may have had no information on which to base their opinions.

**Study 2**

Educators are increasingly using course Web sites to supplement information presented in the classroom (National Education Association, 2001). Study 2 examined the efficacy of an optional Web site for an introductory psychology course. In addition to examining the association between use of the Web site and course grades, we sought to determine whether use of the Web site was associated with students’ experiences with and attitudes toward computers. Our hypotheses for Study 2 were (a) students who used the Web site would outperform students who did not use the Web site when ACT scores were statistically controlled, (b) Web users would prefer the use of the Internet for course-related activities more than nonusers, and (c) students who used the Web site would report more non-course-related computer use, higher computer self-efficacy, and fewer computer hassles at the end of the semester.

**Method**

**Participants.** Students (N = 112) were 42 men and 70 women enrolled in introductory psychology. Fifty-three percent of the sample reported being White, non-Hispanic; 37% African American, non-Hispanic; and 10% other. Sixty-two percent were first-year students, 12% were sophomores, 88% were juniors, and 6% were seniors. Participants ranged in age from 17 to 38 with an average age of 20.67 (SD = 4.40). Forty-eight percent of participants reported owning a personal computer. All participants had the opportunity to use a class Web site generated with the use of BlackBoard software.

**Design and procedure.** We compared students who chose to use the Web site (CA group, n = 60) to those who did not use the Web site (traditional group, n = 52). The Web site offered a copy of the instructor’s class notes, a 30-item multiple-choice practice test for each exam, announcements of due dates and upcoming exams, a copy of the syllabus, and study guides for each exam. The instructor provided all participants with paper copies of the syllabus and study guides as well as in-class oral announcements of upcoming due dates and exam dates.

We used the following instruments: a 13-item demographic questionnaire, students’ final grades in the course (scores included 10 multiple-choice exams and 24 assignments), a mean of all exam scores, a 2-item Likert-type scale measuring student liking of Internet activities designed for classes (1 = strongly disagree, 5 = strongly agree), an 8-item Likert-type evaluation of the course Web site (1 = strongly disagree, 5 = strongly agree) completed by participants who had used the Web site by the end of the semester, a 26-item Likert-type assessment of computer self-efficacy (Torkzadeh, Pfughoeft, & Hall, 1999), a 37-item assessment of computer hassles (scored on a 4-point scale rating the severity of the problem; Hudiburg, 1989, 1995), and a 21-item computer use questionnaire (scored on a 5-point scale rating the frequency of use; Panero, Lane, & Napier, 1997). We administered the last three scales twice, once at the beginning and once at the end of the semester.

**Results**

Students in the CA group reported accessing the Web site a mean of 13.03 times (SD = 8.34) by the end of the semester. A test showed that the CA group and the traditional group were not significantly different in ACT scores (CA group: M = 18.75, SD = 2.83; traditional group: M = 18.12, SD = 2.66), t(110) = −1.21, p = .23, η² = .01, observed power = .22.

Hypothesis 1 predicted course Web site use would be associated with improved student grades. Results of ANCOVAs indicated that the two groups did not significantly differ in performance when ACT scores were statistically controlled. ANCOVA showed similar mean exam scores for the CA group (M = 75.10, SD = 13.12) and the traditional group (M = 71.22, SD = 16.61), F(1, 105) = .61, p = .44, η² = .01,
observed power = .12. Analysis of final course grades showed the same pattern (CA group: $M = 77.30, SD = 14.02$; traditional group: $M = 75.81, SD = 12.51$), $F(1, 104) = .04, p = .84$, $\eta^2 = .00$, observed power = .06.

As Hypothesis 2 predicted, the CA group ($M = 4.13, SD = .66$) reported liking Internet activities more than the traditional group ($M = 3.84, SD = .69$), $F(1, 111) = 4.93, p < .05$. The CA group also had a marginally positive response to the course Web site ($M = 3.61, SD = .64$). Thus, even though use of the Web site was not associated with higher grades, CA participants rated the Web site somewhat positively and liked the use of the Internet for course-related activities more than traditional participants did.

The third hypothesis predicted an association between Web site use and self-reported experiences with computers. More specifically, we expected that students who used the Web site would show higher scores on measures of computer use and computer self-efficacy and lower scores on computer hassles at the end of the semester. To control for potential differences between the CA and traditional groups, we used ACT scores and presemester scores (i.e., measures of computer use, computer self-efficacy, and computer hassles at the beginning of the course) as covariates in ANCOVAs. The two groups did not differ significantly in computer use, $F(1, 79) = .33, p = .84$, $\eta^2 = .00$, observed power = .05; or computer self-efficacy, $F(1, 79) = .97, p = .19, \eta^2 = .02$, observed power = .26. Students in the CA group ($M = 1.01, SD = .69$) reported less severe computer hassles than those in the traditional group ($M = 1.58, SD = .71$) at the end of the semester, $F(1, 78) = 11.09, p < .05$. Thus, Hypothesis 3 was partially supported.

General Discussion

Comparable to previous research (Aberson et al., 2000; Brothen, 1997; Graham, 1997; Newcomb et al., 1998; Plous, 2000; Sherman, 1998; Varnhagen et al., 1997), we found that students who were exposed to CA instruction rated it somewhat positively (Study 2), preferred it to other techniques (Study 1), or liked it more than students who were not exposed to it (Studies 1 and 2). Generally, our findings suggest that instructors could use classroom technology to attract students and enhance student enjoyment of courses. In addition, Study 2 found that students in the CA condition experienced less severe computer hassles than those in the traditional condition even when presemester hassles were statistically controlled. This result indicates that classroom computer technology may facilitate the computer skills of students. However, it is possible that students who chose to use the Web site also participated in other computer activities that led to a reduction in computer hassles.

As educators, our most important question in this investigation was whether revising our curriculum to include technology enhanced student learning. We were disappointed to find that neither the use of computer-generated slides (Study 1) nor an optional Web site (Study 2) resulted in improvements in student performance on course exams or final course grades. Our findings are consistent with some previous research (Kazmerski & Blasko, 1999; Stoloff, 1995; Welsh & Null, 1991), but are in contrast to other studies (Erwin & Rieppi, 1991; Forsyth & Archer, 1997; Worthington et al., 1996). Of these studies, some have been confounded by the use of different instructors for CA and traditional conditions (Erwin & Rieppi, 1999; Welsh & Null, 1991), by having no control group (Forsyth & Archer, 1997), or by mandating additional instruction time for CA conditions (Worthington et al., 1996). Previous studies that controlled for these problems (Kazmerski & Blasko, 1999; Stoloff, 1995) and this study, which used a quasi-experimental approach, found no statistically significant effects on student exam scores. These results suggest that future tests of CA instruction in psychology courses should strive to increase the discrepancy between the control and experimental groups by strengthening the CA intervention. Alternatively, future research could focus on identifying specific student groups (e.g., highly visual learners; Smith & Woody, 2000) to determine whether certain types of students benefit more than others from CA instruction.

Our studies were limited by several factors. One limitation was the lack of random assignment to conditions. Even though we statistically controlled for group differences in ACT scores, there may have been other differences between the groups. Second, participants were students at an open-admissions, historically Black university. Hamilton (2001) noted that historically Black university students are less likely to have access to computers when compared to students at other universities. Less than half of the students in Study 2 indicated that they owned a computer. Perhaps students with more access to computers would have yielded different results. Finally, for both our studies, estimates of statistical power and effect sizes were small. The combination of low statistical power and small effect sizes may be the reason for nonsignificant results. Although the sensitivity of our statistical tests would have been enhanced by larger sample sizes, the effect sizes would remain small.

In an effort to upgrade technology in psychology courses, the administration at our university awarded us high-end computer resources and intensive training on their use. Such awards were relatively rare on our campus at the time of this investigation. We conducted these studies to demonstrate that such investments improve student learning. We found that students liked our technology-enhanced classes, but the students did not earn better grades. The fact that students like CA instruction may be valuable to administrators and educators who are attempting to recruit and satisfy students. However, these benefits are moderated by several factors, including the cost of equipment purchase and maintenance, the cost of faculty training, and the time expenditure necessary for developing new classroom materials. To date, research has not yet established the types of CA teaching interventions that are effective in improving student performance in psychology courses. Therefore, we recommend that administrators conduct thorough cost–benefit analyses before devoting extensive resources to instructional technology.

References


**Notes**

1. We thank Walter C. Johnson, Carlos A. Roberts, and Karen Jung for their helpful comments in the preparation of this article.

2. Send correspondence to Kurt A. DeBord, Division of Social and Behavioral Sciences, Lincoln University, Jefferson City, MO 65102–0029; e-mail: debordkurt@yahoo.com.
Computer-administered exams offer many advantages, but instructors may be reluctant to use them due to concerns that computer anxiety may increase student test anxiety. Introductory psychology students (N = 265) completed surveys prior to their first exam about their anxiety related to the upcoming exam, computers in general, and taking exams on the computer. One group of students took traditional paper-and-pencil exams and the other group took computer-administered exams. We found no differences between the groups for exam anxiety or general computer anxiety, but the traditional group reported more anxiety about taking an exam on computer. We recommend strategies for relieving students' anxiety, such as in-class demonstrations of the technology.

Many introductory psychology textbooks offer online testing as part of the instructor's resources. Testing online offers several advantages to both students and instructors. Students receive quick feedback on their exam score and often which questions they answered correctly. Instructors spend less time grading, and eliminating photocopying saves money.

However, online testing has its disadvantages. New technology invites new ways for students to cheat; therefore security concerns must be addressed by the testing program. Some instructors may be unwilling to invest the time needed to learn a new technology. Finally, computer testing may cause students increased anxiety, above the normal amounts of test anxiety. Our study addressed this last concern by surveying introductory psychology students preparing to take their first exam for both test anxiety and computer anxiety.

Test Anxiety

Test anxiety is concern about negative evaluation that students experience before and during a test. The anxiety may take the form of worry—unwanted, negative thoughts about one's performance, or emotionality—physiological symptoms such as increased heart rate and sweaty palms (Hembree, 1988; Powers, 2001). Hembree's meta-analysis of 562 studies found negative correlations between test anxiety and IQ, GPA, course grades, and achievement scores in reading, math, natural sciences, and other subjects. Hembree concluded that test anxiety is not only related to poor performance, but is the cause of it because interventions that lowered test anxiety resulted in higher achievement. Furthermore, women reported more test anxiety than men, although their performance levels were the same.

Computer Anxiety

Chua, Chen, and Wong (1999) defined computer anxiety as a fear experienced when using a computer or thinking about using a computer. Their meta-analysis of studies published since 1990 showed that computer experience was negatively related to computer anxiety. However, more recent research suggests that undergraduates' familiarity with and access to computers are not related to computer anxiety (McIlroy, Bunting, Tierney, & Gordon, 2001). Yet, students who had a positive initial experience in computing were less anxious than those who had a negative first experience, and students taught by a confident and competent computer instructor (as perceived by the students) had more positive attitudes toward computers than students whose instructors lacked these qualities.

Testing on Computers

The relation of test anxiety to computer testing is not clear, with some research reporting that computer anxiety was negatively related to test performance (e.g., Johnson & Johnson, 1981), and other research finding no relation between computer anxiety, type of test—either computer or paper-and-pencil—and performance (e.g., Wise, Barnes, Harvey, & Plake, 1989). Vispoel, Rocklin, and Wang (1994) found that students generally preferred computerized testing to paper-and-pencil formats, but disliked tests that did not allow for item skipping or review.

We examined undergraduates' self-reports about test and computer anxiety when faced with one of two testing situations—a traditional paper-and-pencil multiple-choice test or a computer-administered test. Before the first exam, we surveyed general psychology students about their anxiety toward the upcoming test, computers in general, and taking tests on a computer.

Method

Participants

A total of 265 students participated. Students received course credit for their participation. Two sections of introductory psychology (163 students; 53 men, 110 women)
taught by two different instructors used computerized testing procedures. Instructors administered the computer exams during the scheduled class time in campus computer laboratories with the instructor or an assistant present. The computer group included 104 freshmen, 38 sophomores, 16 juniors, and 5 seniors. Two sections of introductory psychology (102 students; 47 men, 55 women) taught by different instructors used traditional paper-and-pencil exams scored by machine. The traditional group included 67 freshmen, 23 sophomores, 8 juniors, and 4 seniors. The two groups were similar in age (computer: $M = 21.01$, $SD = 4.45$; traditional: $M = 21.46$, $SD = 5.15$). We did not inform the students of the format of the exams before they enrolled in their section of the course.

Instrument

We adapted the State Anxiety in Computing Situations portion of the Computer Anxiety and Learning Measure (CALM; McInerney, Marsh, & McInerney, 1999) for this study. The State Anxiety subscale consists of 20 items, which represent four factors: worry, happiness, physiological symptoms, and distractibility. We presented these items in three yes–no checklists. The students indicated which feelings (e.g., worried, threatened, comfortable) or symptoms (e.g., dry mouth, sweaty palms) they had in relation to three situations: the upcoming exam, computer use in general, and taking the exam on a computer. The traditional group imagined that they would take the exam on the computer for the last list.

Procedure

We distributed the surveys during a regular class meeting 1 week before the first exam. All of the exams were multiple-choice. Approximately 90% of the students chose to participate. Prior to the survey, we told all of the students in which format (computer or paper and pencil) their exams would be administered.

Results

We tabulated anxiety scores by adding the number of yes responses for the worry, physiological symptoms, and distractibility items and the number of no responses for the happiness items (range = 0 to 20) for each of the situation checklists. A 2 (group: computer or traditional) × 2 (male or female) × 3 (situation: exam, computer, exam on computer) repeated-measures ANOVA on the anxiety scores found several significant effects ($p < .05$). The significant Group × Situation interaction answered the question of whether taking the test on computer was related to greater anxiety than the traditional test administration, $F(2, 522) = 8.41, p < .001$, partial $\eta^2 = .031$. Post hoc Tukey tests found that the thought of taking the test on computer was more anxiety provoking for the traditional group ($M = 7.10$, $SD = 5.34$) than for the computer group ($M = 5.45$, $SD = 5.08$), $p < .01$. There were no differences between the groups on anxiety toward the exam itself or toward computers in general (see Table 1 for means). For the computer group, students reported more anxiety for the exam itself than taking the exam on computer, $p < .01$, but the traditional group showed no difference between these situations. For both groups, students reported significantly less anxiety for computers than the exam or taking the exam on a computer, $p < .01$.

The Sex × Situation interaction was significant as well, $F(2, 522) = 3.75, p = .02$, partial $\eta^2 = .014$. Post hoc Tukey tests found that women ($M = 9.09$, $SD = 4.93$) reported more anxiety than men ($M = 6.81$, $SD = 4.97$) about the upcoming exam, $p < .01$. However, there were no sex differences in anxiety regarding computers in general or taking the exam on computer (see Table 2 for means).

As expected from the interactions, there were also significant main effects for situation, $F(2, 522) = 105.91, p < .001$, partial $\eta^2 = .289$, and sex, $F(1, 261) = 9.71, p = .002$, partial $\eta^2 = .014$. We did not inform the students of the format of the exams before they enrolled in their section of the course.

<table>
<thead>
<tr>
<th>Table 1. Mean Anxiety Levels for Group × Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Computer</td>
</tr>
<tr>
<td>Traditional</td>
</tr>
</tbody>
</table>

Note. Possible scores range from 0 to 20, with 20 representing the highest anxiety. Means in the same row that do not share subscripts differ at $p < .05$ in post hoc Tukey tests.

<table>
<thead>
<tr>
<th>Table 2. Mean Anxiety Levels for Sex × Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>

Note. Possible scores range from 0 to 20, with 20 representing the highest anxiety. Means in the same row that do not share subscripts differ at $p < .05$ in post hoc Tukey tests.
Neither the main effect for group nor the Group × Sex × Situation interaction were significant.

Discussion

Our results indicate that, when initially confronted with the thought of computer-administered exams, some students are apprehensive. The traditional group reported more anxiety about the thought of computer testing than those students who actually took the tests on computers. The traditional group was no more anxious than the computer group about the exam or computers in general—the anxiety related specifically to the format of the exam. The computer group knew the format of their exams from the first day of class, whereas it is possible many students in the traditional group never considered their feelings toward taking their psychology exams on computer until they completed the survey. The extra time the computer group had to become accustomed to the idea of computer testing may account for their lower levels of anxiety toward computer exams compared to the exam itself or compared to the traditional group’s level of anxiety towards computer exams. Our results also support the sex differences in test anxiety found by Hembree (1988) and the lack of sex differences in computer anxiety found by Chua et al. (1999).

The apprehension of both groups toward computer testing may be due to the unfamiliarity of the task and past test-taking experience. General psychology often represents one of students’ first college classes, and the likelihood that this is the first course to have computer-administered exams may add to the doubts and fears of beginning college work. On the other hand, as computers become an integrated part of the world and classroom, students may come to accept computer exams as an extension of the larger changes in their world. The routine administration of some standardized exams such as the SAT and the GRE on computers may hasten such acceptance.

Students’ initial experiences in computing situations are important for reducing anxiety (McIlroy et al., 2001). Instructors who plan to use computer-administered exams should be aware of the potential for anxiety, doubts, fears, and concerns among their students and should attempt to alleviate them prior to the first exam by such measures as general discussion of the issues or demonstrations and hands-on trial experience with computer-administered exams (e.g., showing projected demonstrations of a hypothetical exam experience, placing practice exams on the computer). Instructors may also permit those students with high levels of anxiety to opt for paper-and-pencil exams when feasible. Such students may later feel confident enough to take computer-administered exams as they observe their fellow students successfully completing computerized exams (we prefer to have such students take the exam in the same computer laboratory at the same time as those who take the exam by computer).

However, further study of computer testing is warranted. For example, a study of not only anxiety measures but test results over the course of the class may determine whether student performance is different with computer-administered as compared to traditional examination methods. We did not consider student performance in this study because of the lack of consistency inherent in comparing four different instructors using four different exams. Information about students’ prior experiences with computerized exams might also be examined, particularly as the use of computers in education increases not only at the college level, but also at the secondary education level.

Computers provide a potentially positive match to budget and personnel constraints that many universities face. Particularly in the case of large sections of courses such as general psychology, the use of computer-administered exams seems a viable solution to some of these practical problems. Provided that possible initial student anxiety is confronted, there are benefits for the university, faculty, and students with the use of computer-administered exams.

References


Note

Send correspondence to Carolyn A. Schult, Department of Psychology, Indiana University South Bend, P. O. Box 7111, South Bend, IN 46634–7111; e-mail: cschult@iusb.edu.
Teaching of Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/htop20

Employing Computer-Administered Exams in General Psychology: Student Anxiety and Expectations
Carolyn A. Schult & John L. McIntosh

Available online: 20 Nov 2009

To cite this article: Carolyn A. Schult & John L. McIntosh (2004): Employing Computer-Administered Exams in General Psychology: Student Anxiety and Expectations, Teaching of Psychology, 31:3, 209-211

To link to this article: http://dx.doi.org/10.1207/s15328023top3103_7

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Using Group Web Page and Video Clip Creation Exercises in Introductory Psychology Courses

Terry F. Pettijohn II and Elizabeth G. Perelli
Mercyhurst College

Students (N = 66) enrolled in an introduction to psychology courses at a small, private college created group Web pages and video clips and then shared these creations with classmates as part of group project on major topic areas covered in class. As predicted, students scored significantly higher on comprehensive final exam multiple-choice questions pertaining to their group project topic area compared to questions in other topic areas. Furthermore, students reported that they enjoyed working with other students in their group, enjoyed completing the project requirements, felt more comfortable using technology, and believed at the end of the term that the group project helped them learn about psychology.

Professors strive to facilitate student learning and excite students about course content using various instructional methods. Technology and the Internet offer exciting potential to generate student interest and interactive, collaborative learning experiences. Although research has shown the benefits of group projects on undergraduate psychology course outcomes (e.g., Millis, 2001), the role of technology in this relation has not been explored.

We developed a group project to assess this combination and evaluate the value of assigning group projects with technology components. The assignment was worth approximately 5% of the final course grade. Groups of 3 or 4 students explored a unit within an introduction to psychology course (biopsychology, developmental, sensation/perception, learning, cognition, motivation/emotion, personality, clinical, or social psychology). Part 1 of the group project required students to create a group Web page to learn more about their specific topic area and teach other students about their particular unit. The content was to include a general introduction and overview of the unit, at least 10 substantial information Web site links with brief descriptions that further clarified the unit, and at least two interactive Web exercises (i.e., quizzes, demonstrations, dynamic images). Students used Microsoft Word® to create their group Web pages, and the first author (the course instructor) posted the work on the course Web site. The instructor led a brief instructional session on using Word to create Web pages at the beginning of the term. This orientation was necessary because only 2 of 67 initial students indicated prior experience making Web sites.

Part 2 of the group project required students to create a group video clip that expanded and demonstrated their knowledge of their same assigned topic area and to share this information with the class. Students produced a short (approximately 3 min) digital video segment that provided a real-life example or application of the information covered from their assigned unit. Students used a digital camcorder to record these clips. Some examples included skits of how individuals recognize emotion; interviews with other students showing how memory and motivation affect students during exam time; and examples of psychological assessment tools, such as the Rorschach Ink Blot Test, to measure personality. Groups shared these short video clips in a Microsoft PowerPoint® presentation at the end of the term, as part of a class review session for the comprehensive final exam.

The group Web page project and group video clip project allowed students the opportunity to work collaboratively, apply their knowledge, think creatively, use technology, and share information about psychology. Due to the interactive nature of group projects and the in-depth nature of the project, we hypothesized that students would score higher on their respective group topic area compared to other areas on a comprehensive final exam. We also hypothesized that completion of the group project would facilitate learning, make psychology more enjoyable, and help students become more comfortable using technology.

Method

Participants

Sixty-six students (50 women, 16 men) enrolled in two concurrent sections of an introduction to psychology at a small, private college in the northeastern United States participated in this study. Students were traditional-aged and predominantly White, representing the typical student enrolled at this college. I (TFP) randomly assigned students in each section to groups of 3 or 4 and assigned the groups one of nine course topic areas coinciding with course instruction.

Materials and Procedure

To assess the effectiveness of the group projects, I administered a comprehensive multiple-choice final exam to each class at the end of the term. I included questions from course material divided into topic areas that coincided with the topic
areas assigned to student groups. Tests included between 8 and 12 questions to assess each of the topic areas. I created the final exam prior to the presentation of group projects and therefore did not design the exam to specifically test group project information. There was some general overlap between group presentation topics and exam questions, but this overlap was distributed evenly among group areas. More important, the information reported by groups was shared with the entire class, not just between the group and the instructor.

In addition, we created a survey about the use of technology and how these group projects affected the student learning experience and distributed the questionnaire after the groups completed their projects. The survey asked students to indicate yes or no to the following questions: Did you enjoy working with the other students in your group? Did you enjoy completing the project requirements? Did you feel more comfortable using technology after completing this project? Did the group project help you learn psychology? The survey had additional space for students to explain why or why not and to make general comments about the experience.

Results

As predicted, students scored significantly higher on their respective content area compared to other content areas on the final exam, \( t(65) = 3.00, p = .004 \) (87.1% vs. 80.0% correct), \( d = .37 \). The majority of the students also indicated that they enjoyed working with other students in the group, \( \chi^2(1, N = 59) = 51.27, p < .001 \) (96.6% = Yes); they enjoyed completing the project requirements, \( \chi^2(1, N = 59) = 47.61, p < .001 \) (94.9% = Yes); they felt more comfortable using technology after completing the project, \( \chi^2(1, N = 59) = 18.46, p < .001 \) (78.0% = Yes); and that the group project helped them learn psychology, \( \chi^2(1, N = 59) = 47.61, p < .001 \) (94.9% = Yes). All of those who indicated they did not feel more comfortable using technology after completing the project explained that they were already comfortable using technology before conducting the project.

Discussion

This project demonstrates how professors can incorporate technology-rich group exercises to enrich the learning experience in introductory psychology courses. Not only did this project lead to increased learning of a particular area of psychology, it was also an experience students enjoyed and an exercise that increased their comfort level using technology. Almost all students indicated they had no previous experience making Web pages at the start of the term, and few students had used digital camcorders, but by the end of the term, each group had produced quality projects. The effort involved in making the Web pages and video clips was quite thoughtful and creative. Students enjoyed working together, using technology, and learning about psychology. Apart from the surveys and final exam scores, students also mentioned their positive experiences with the project informally outside of class and on course evaluations.

Does this group project increase depth of material understanding at the expense of breadth of material understanding? We do not believe so. Although we did not have another class without group projects for comparison purposes, we did consider the results of a final exam from a previous section of introductory psychology taught by the first author. Although actual course content was not identical and some of the final exam questions were slightly different, final exam scores from these two sections were not significantly different. The similar final exam scores suggest that the level of material understanding at the end of the course for the group project classes was most likely not confined to one single area. Additional studies may investigate the depth versus breadth issue in the future.

There are several potential refinements to the project that may be addressed in future research. Redesigning the assessment tool to include Likert-type scales to evaluate student perceptions would provide a more sensitive indication of the effectiveness of the group project. Additional questions about the amount of time students invested in the project or how much effort each group member contributed may also be useful information. If a single group project can have these effects, perhaps requiring additional, smaller projects using technology for each content area throughout the term may also increase learning in multiple areas.

Although this group project was a worthwhile experience, it did require additional effort and time on the part of the instructor. The first author was familiar with developing Web pages and using digital video prior to incorporating this requirement, and this familiarity was necessary to answer questions and troubleshoot throughout the term. Lack of this knowledge would require additional preparation time and could make the project unfeasible. For each section, the instructor spent an estimated 30 min in class leading an informational session on how to make Web pages, 1 hr transferring completed group Web pages to the Internet, and 2 to 3 hr transferring the digital video from the camcorder onto a computer and converting the footage to a digital file to incorporate into a PowerPoint presentation. The instructor also spent some extra time with individual students, grading contributions, and reserving equipment. The unavailability of resources and larger class sizes may be an obstacle for some professors.

Student effort and time involved with the project is more difficult to estimate, as some groups certainly spent more time and effort on their creations than others. A rough estimation of time invested in these group projects would be 3 to 5 hr for the Web page and 3 to 5 hr for the video clip creation. Some of the groups met together initially, developed a plan, divided the tasks among members, completed some individual work, and met again to put all the individual pieces together. Students indicated some minor problems getting together with their groups outside of class due to work, sports, and other conflicting activities. In all, this was a positive exercise that facilitated learning, excitement for psychology, and comfort using technology.

Reference


Notes

1. This research study was partially supported by a Preparing Tomorrow’s Teachers to Use Technology (PT3) Grant awarded to Mercyhurst College.
Group Differences in Academic Achievement: Service Learning in a Child Psychology Course

Kari Knutson Miller and Shu-Chen Yen
Department of Child and Adolescent Studies
California State University, Fullerton

We analyzed the relation between service learning and academic achievement and sought to determine whether characteristics of the service experience mediate academic outcomes. We compared outcomes associated with nonservice-learning, indirect service-learning, and direct service-learning conditions. Students participating in direct service-learning experiences achieved a greater mastery of course learning goals than students participating in indirect service-learning activities. Students participating in direct service-learning experiences also scored higher on the final exam than nonservice learners.

The purpose of this investigation was to examine the impact of service learning on academic outcomes. Although the literature generally notes a positive relation between service learning and academic achievement, questions remain due to measurement or design issues such as student self-report (Hardy & Schaen, 2000), self-selection of service-learning or nonservice-learning options (Reeb, Sammon, & Isackson, 1999), and instructor effects (Marcus, Howard, & King, 1993).

Shastri (2001) and Strage (2000) addressed several of these concerns in studies of performance outcomes associated with service-learning experiences in educational psychology and child development courses, respectively. Shastri reported nonsignificant group differences on achievement as measured by exams. Strage found group differences were significant on second and final exams; further analysis of the second exam revealed that service learners achieved higher scores on essays, but not multiple-choice components. As instructors commonly use exams that include multiple-choice questions to measure academic outcomes, specific characteristics related to the service-learning experience that promote achievement in this context need further exploration.

This investigation is an extension of Knutson Miller, Yen, and Merino (2002); they examined academic outcomes related to non- and indirect service-learning experiences in a child psychology course. Indirect service-learning experiences were those in which students learned about a community, applied course knowledge to create a service or product designed to meet community needs, and analyzed course content through this application (Connor-Linton, 1995). In the context studied by Knutson Miller et al. (2002), indirect service-learning participants developed resources based on needs articulated by tutors in the America Reads and Counts (ARC) Program and disseminated these resources at tutor training sessions. Participants submitted reflections connecting this experience with course learning goals at the conclusion of the semester. Results indicated between-group differences on measures of academic achievement, including exams, were not significant. The authors suggested two potential modifications for future service-learning integrations. First, they recommended that instructors consider the potential impact of direct versus indirect service-learning experiences on the achievement of course learning goals. Second, the authors suggested that instructors design prompts for reflection that require participants to link service experiences and course themes on a regular, systematic basis. The purpose of this study was to compare academic outcomes across three conditions. In addition to the conditions studied by Knutson Miller et al., we considered the impact of direct service-learning experiences incorporating structured reflection prompts.

Method

Participants

Participants included 266 students enrolled in an upper division undergraduate child psychology course during three subsequent semesters (Semester 1, n = 103; Semester 2, n = 91; and Semester 3, n = 72). The majority of the participants were women (n = 259), and all were junior- and senior-level child and adolescent development majors. For the purpose of this investigation, we designated Semester 1 as nonservice learning, Semester 2 as indirect service learning, and Semester 3 as direct service learning. Participants were unaware of these designations prior to the first day of class.

Course Description

Course content, exams, and instructors were consistent across conditions. Noncumulative midterm and final exam questions required analysis of contextual situations and application of course themes in a multiple-choice format. All students also completed an applied course project that required identification of a theme related to development in middle childhood, review of relevant scholarly literature, and identification of practice implications in written format (e.g., newsletters, handbooks, resource guides). Project assessment was based on a standardized grading rubric.

Indirect Service-Learning Integration

Indirect service-learning participants determined the content of their course projects by attending to needs articulated by tutors from the ARC Program. The investigators designed a survey to identify challenges tutors encountered in their interactions with elementary school children. After participants reviewed tutor needs, they identified those related to course themes. Participants then prepared their applied course pro-
Past research on the effects of presentation software has relied on small samples and experienced instructors. My research used a quasi-experimental pretest–posttest design to evaluate the impact of PowerPoint® on student learning, satisfaction, and engagement in an introductory psychology course taught by graduate student instructors. Results showed several main effects of instructor but virtually no effects of PowerPoint, although there was a significant instructor by PowerPoint interaction on perceived learning and interest in psychology. PowerPoint reduced perceived learning for one instructor, but increased interest in psychology for another. The results are a reminder that good teaching depends more on the instructor than the technology.

Research suggests that using presentation software (e.g., Keynote or PowerPoint®) in teaching may have a positive impact on college student learning (Kulik, Kulik, & Cohen, 1980). Presentation software allows instructors to present text and images, including photographs, in color; to animate text and graphics; and to include audio and video files. The software is inexpensive and relatively easy to implement, making it a particularly important tool in maximizing the number of students who may benefit from the integration of technology into teaching.

However, relatively little research has examined the effects of presentation software on student outcomes. One recent study (DeBord, Aruguete, & Muhlig, 2004) compared students in one section of introductory psychology, in which the instructor delivered lectures using PowerPoint, to students in a second section, in which the same instructor delivered lectures using overhead transparencies. The authors found no differences in final exam scores. However, some students do seem to benefit from the use of visual displays in teaching. Students with a highly visual learning style had better exam scores when in a multimedia class (Smith & Woody, 2000), and experienced and novice students learned more from instructional content that included animations than from purely text-based instructional content (ChanLin, 1998).

Although these studies have improved on earlier research by including control groups and controlling for instructor effects, they have still relied on small samples (usually comparing one section of a course to another). Furthermore, in these studies (DeBord et al., 2004; Smith & Woody, 2000), the instructors have tended to be the study authors, who implicitly or explicitly support the use of such technology and are presumably experienced in using it. At many large universities, however, inexperienced graduate students teach introductory courses. In addition, anecdotal evidence suggests that the use of presentation software is increasingly expected or required of college-level instructors. However, surveys of students suggest that students report frustration with their instructors’ use of presentation software and, in some cases, prefer less technology in the classroom, rather than more (Young 2004a, 2004b).

My purpose was to empirically test whether the use of presentation software is preferred by college students and leads to better student outcomes. Using eight sections of introductory psychology taught by four graduate student instructors, I used a quasi-experimental pretest–posttest design to evaluate the impact of using PowerPoint on student learning, satisfaction, and engagement in an introductory psychology course.

Method

Participants

Undergraduates enrolled in one of eight sections of introductory psychology at Texas Tech University par-
ticipated in the study. Four different graduate student instructors who had been assigned to teach Introductory Psychology and who subsequently volunteered to assist with the study in exchange for a monetary bonus taught these sections. Across these eight sections, there were 263 students in the initial sample, during the first week of class, and 181 students remaining in the final sample in the last week of class. Demographic data were missing for approximately one third of the sample, but remaining demographic data indicated the sample was composed of predominantly non-Hispanic White (81.8%) first-year (51.1%) students, evenly divided across gender (52.8% women). Of the instructors, three were women and one was a man. One instructor was Hispanic, one was African American, and the remaining two were non-Hispanic White. One had not taught introductory psychology before, one had taught the course in one previous semester, and two had taught for three or more semesters. One had experience teaching both with and without PowerPoint. The remaining three had never taught using PowerPoint.

Measures and Procedure

To control for instructor effects, course instructors who volunteered to assist with the study each taught one section of introductory psychology using PowerPoint and another section without using PowerPoint. All instructors taught both of their sections on the same day; two of the instructors used PowerPoint in their first section, and two used PowerPoint in their second section of the day. The instructors received basic training in the use of PowerPoint in delivering lectures by the study author prior to the start of the semester. In the PowerPoint sections, instructors used PowerPoint to present all lecture content. All instructors used the same textbook and taught the same basic course content. However, individual instructors had flexibility in other aspects of the course, such as testing and ordering of the course content. I asked each instructor to teach both of his or her sections the same way, with the exception of how they delivered lectures (with or without PowerPoint).

During one of the first two class sessions of the semester, instructors administered a 44-item questionnaire to all students in attendance. In addition to basic demographic questions, there were questions about how much students expected to enjoy the class on a scale from 1 (not at all) to 5 (very much), how much they expected to learn in the class compared to other courses they were taking that semester on a scale from 1 (much less than average) to 5 (much more than average), how interesting they expected psychology to be on a scale from 1 (not at all) to 5 (very much), and how likely they were to take additional psychology courses on a scale from 1 (not at all likely) to 5 (surely take 2 or more courses). In addition, to measure baseline knowledge of course content, students in all sections completed a 35-item multiple-choice test, comprising questions randomly selected from the test banks of two introductory psychology textbooks. During the last week of class, students again completed the same attitudinal questions, with tenses modified as needed (e.g., “How much did you like this class?”). In addition, to obtain an objective measure of student learning, students in all eight sections completed a second 35-item multiple-choice test covering course content.

Results

Preliminary Analyses

To examine whether any of the eight sections differed from each other on any variable at the beginning of the semester, I conducted a 2 (PowerPoint: yes, no) × 4 (Instructor: A, B, C, D) MANOVA. Results of the omnibus F tests showed that the eight sections did not differ from each other on any variable at the beginning of the semester (all Wilks’s λ ≤ .99, Fs ≤ .93, ps > .50). Therefore, the following analyses focus only on end of semester scores (see Table 1). A chi-square test revealed significant differences in the grade distributions for the four instructors; therefore, I used final grades as a covariate in all of the following analyses. A one-way MANOVA indicated no main effect of practice, Wilks’s λ = .97, F(5, 128) = .76, p > .58, indicating that instructors did not have better outcomes for the second section they taught each day, regardless of teaching modality.

Tests of Hypotheses

To examine whether there were effects of instructor or PowerPoint on end-of-semester outcomes after controlling for final grades, I regressed each of the dependent variables on final grades and then conducted a 2 (PowerPoint: yes, no) × 4 (Instructor: A, B, C, D) MANOVA on the saved residuals. I used this procedure instead of entering final grades as a covariate in a MANCOVA to allow post-hoc testing of obtained differences.

Results of the omnibus F test revealed no main effects of PowerPoint, Wilks’s λ = .98, F(5, 168) = .85, p
Scores on the objective learning measure, which contained 35 items, indicate that PowerPoint accounted for less than 3% of the variance in the dependent variables; presence or absence of PowerPoint had no effect on students’ liking for the course, interest in psychology, intentions to take additional psychology courses, perceived learning, or objective learning. However, there was a significant main effect of instructor, Wilks’s \( \lambda = .80 \), \( F(15, 464) = 2.54, p < .001 \), \( \eta^2_p = .07 \), as well as a significant instructor by PowerPoint interaction, Wilks’s \( \lambda = .82 \), \( F(15, 464) = 2.30, p < .01 \), \( \eta^2_p = .06 \).

Univariate tests revealed significant effects of instructor on students’ liking for the course, \( F(3, 172) = 4.64, p < .01 \), \( \eta^2_p = .08 \); perceived learning, \( F(3, 172) = 6.29, p < .001 \), \( \eta^2_p = .10 \); and objective learning, \( F(3, 172) = 2.99, p < .05 \), \( \eta^2_p = .05 \), controlling for final grade. Post-hoc Tukey’s tests indicated that students in instructor D’s classes liked the course more than students in instructor A’s classes and believed they learned more than students in any of the other instructors’ classes. Despite this perception, however, the significant effect on the objective measure of learning indicated that students in instructor B’s classes actually learned more (raw \( M = 17.33 \), SD = 4.50) than students in Instructor D’s classes (raw \( M = 15.70 \), SD = 4.68). However, the main effect of perceived learning was qualified by a significant instructor by PowerPoint interaction, \( F(3, 172) = 4.14, p < .01 \), \( \eta^2_p = .07 \). Post-hoc \( t \) tests (family-wise \( \alpha = .01 \)) revealed that despite no differences in scores on the objective learning measure, Instructor C’s students believed they learned significantly less when the instructor used PowerPoint than when the instructor did not; PowerPoint had no effect on perceived learning for the other three instructors. Finally, there was also a significant instructor by PowerPoint interaction on students’ interest in psychology at the end of the semester, \( F(3, 172) = 4.21, p < .01 \), \( \eta^2_p = .07 \). Post-hoc \( t \) tests (family-wise \( \alpha = .01 \)) revealed that Instructor B’s students were more interested in psychology when the instructor used PowerPoint than when the instructor did not, whereas PowerPoint had no effect on interest in psychology for the other three instructors.

**Discussion**

The results suggest that introductory psychology instructors should use PowerPoint with caution. PowerPoint had no effect on how much students liked their introductory psychology course, their interest in psychology, intentions to take additional psychology courses, or objective or perceived learning. Although

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Teaching Modality</th>
<th>Like M</th>
<th>Like SD</th>
<th>Interest M</th>
<th>Interest SD</th>
<th>Future M</th>
<th>Future SD</th>
<th>Perceived Learning M</th>
<th>Perceived Learning SD</th>
<th>Objective Learning M</th>
<th>Objective Learning SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>No PPT</td>
<td>2.08</td>
<td>.86</td>
<td>2.69</td>
<td>0.63</td>
<td>0.77</td>
<td>0.73</td>
<td>2.31</td>
<td>.95</td>
<td>14.15</td>
<td>4.88</td>
</tr>
<tr>
<td></td>
<td>PPT</td>
<td>1.96</td>
<td>.65</td>
<td>2.22</td>
<td>0.93</td>
<td>0.67</td>
<td>0.92</td>
<td>1.85</td>
<td>.82</td>
<td>15.44</td>
<td>4.72</td>
</tr>
<tr>
<td>B</td>
<td>No PPT</td>
<td>2.27</td>
<td>.70</td>
<td>2.20</td>
<td>0.56</td>
<td>0.87</td>
<td>0.92</td>
<td>2.07</td>
<td>.59</td>
<td>17.33</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>PPT</td>
<td>2.40</td>
<td>.74</td>
<td>3.00</td>
<td>0.65</td>
<td>1.20</td>
<td>1.52</td>
<td>2.47</td>
<td>.64</td>
<td>17.33</td>
<td>4.75</td>
</tr>
<tr>
<td>C</td>
<td>No PPT</td>
<td>2.54</td>
<td>.84</td>
<td>2.68</td>
<td>1.02</td>
<td>1.26</td>
<td>1.10</td>
<td>2.50</td>
<td>.69</td>
<td>15.54</td>
<td>3.23</td>
</tr>
<tr>
<td></td>
<td>PPT</td>
<td>2.21</td>
<td>.82</td>
<td>2.34</td>
<td>0.97</td>
<td>0.79</td>
<td>1.21</td>
<td>1.93</td>
<td>.75</td>
<td>16.17</td>
<td>3.82</td>
</tr>
<tr>
<td>D</td>
<td>No PPT</td>
<td>2.81</td>
<td>.90</td>
<td>2.78</td>
<td>0.87</td>
<td>0.94</td>
<td>1.13</td>
<td>2.66</td>
<td>.70</td>
<td>15.34</td>
<td>4.23</td>
</tr>
<tr>
<td></td>
<td>PPT</td>
<td>2.68</td>
<td>.78</td>
<td>2.50</td>
<td>0.67</td>
<td>1.64</td>
<td>1.40</td>
<td>2.82</td>
<td>.66</td>
<td>16.23</td>
<td>5.34</td>
</tr>
</tbody>
</table>

**Note.** Students rated how much they liked the class and how interesting they thought psychology was based on a 5-point scale ranging from 1 (not at all) to 5 (very much). Students also rated how much they learned in the class, compared to other courses they were taking that semester, based on a 5-point scale ranging from 1 (much less than average) to 5 (much more than average). Finally, students indicated how likely they were to take additional psychology courses based on another 5-point scale ranging from 1 (not at all likely) to 5 (surely take 2 or more courses). Scores on the objective learning measure, which contained 35 items, indicate the average number of correct items. PPT = PowerPoint.
PowerPoint seems to have facilitated Instructor B’s teaching to a small degree, with students reporting more interest in psychology when this instructor used PowerPoint, PowerPoint actually detracted from Instructor C’s teaching, with students reporting they learned less when this instructor used PowerPoint, despite no actual differences in objective learning. PowerPoint had no effect on perceived learning or interest in psychology for students in the classes taught by the other instructors.

Attitudes toward presentation software tend to be polarized, with some instructors believing its use detracts from their teaching and student learning (see Tufte, 2004), and others believing its use enhances teaching and learning, at least for some students (e.g., Smith & Woody, 2000). By demonstrating that PowerPoint enhanced teaching for one instructor but detracted from teaching for another, these results suggest both positions may be correct. Instructors should carefully consider their teaching style, comfort using technology, and training and experience using presentation software before using it in their teaching. In my study, use of PowerPoint had no effect on objective learning and virtually no effect on students’ attitudes toward the course. Undergraduate students do not necessarily prefer classes that use presentation software. The significant instructor by PowerPoint interactions indicate that although PowerPoint may enhance the teaching of some instructors, it may also detract from the teaching of others, at least in terms of student perceptions and interest in psychology.

Indeed, the most salient finding from this research is an important, but sometimes neglected, point: It is the instructor, not the technology, that is most critical. Instructor effects were stronger than effects of PowerPoint. Instructor D consistently stood out as a better teacher than the other instructors, regardless of the presence or absence of PowerPoint. Teaching experts do not consider use of technology (including PowerPoint) essential for a good lecture (Benjamin, 2002), and technology use is rarely among the characteristics of “master teachers” (and when it is, it is rated as among the least important characteristics by both students and faculty; Buskist, Sikorski, Buckley, & Saville, 2002). My study adds empirical evidence to support this anecdotal evidence. Good instructors are good instructors, regardless of what modality they use.

In the face of emerging instructional technology, both new and experienced instructors may lose sight of the importance of basic teaching skills and be tempted to devote more time to learning the latest technology than to developing or improving these basic skills. As anecdotal student reports seem to indicate (Young, 2004a, 2004b), using presentation software may actually accentuate rather than hide deficits in basic teaching skills. My research suggests that students appreciate and learn from the instructor, not the technology. In the rush to embrace emerging technology, we must remember that whether technology enhances teaching is an empirical question. As educators learn more about what technology is effective when used under what conditions and for which students, they must remember that what is important is the quality of the teaching, not the technology.

References


Notes

1. This research was supported by a Faculty Incentive Grant from the Teaching and Learning Technology Center at Texas Tech University.
Microsoft Producer: A Software Tool for Creating Multimedia PowerPoint® Presentations

Thad R. Leffingwell, David G. Thomas, and William H. Elliott
Oklahoma State University

Microsoft® Producer® is a powerful yet user-friendly PowerPoint companion tool for creating on-demand multimedia presentations. Instructors can easily distribute these presentations via compact disc or streaming media over the Internet. We describe the features of the software, system requirements, and other required hardware. We also describe 3 approaches to using the software to enhance teaching of graduate and undergraduate psychology courses, offer tips on effective use of Producer, list other similar software, and note some shortcomings. Producer can be a useful tool for the teaching of psychology via both traditional and online formats.

Computer-based technologies have revolutionized college-level instruction primarily due to the addition of several new pedagogical techniques that are available. Instructors use the Internet to make class materials available to students and offer entire courses online (ChanLin, Huang, & Chan, 2003), allowing access and flexibility far beyond what was possible 20 years ago. The Internet can also be used for class assignments that require using the power of information-retrieval technologies. Within the classroom, instructors may use computer-assisted tutorials and other individual assignments, techniques that have shown some degree of efficacy in improving student performance (Worthington, Welsh, Archer, Mindes, & Forsyth, 1996). The use of individual response systems (“clicker technology”) in the classroom, whereby each student has a keypad with which he or she can respond anonymously to questions, allows for the immediate tabulation and graphing of data. This active-learning tool increases student interest (Brewster, 1996; Stoloff, 1995) and, in some cases, student performance (Erwin & Rieppi, 2000).

Perhaps the most commonly used computer-based application is presentation software, such as Microsoft PowerPoint. The use of PowerPoint is widespread in classroom instruction, and some data suggest that it enhances student satisfaction and interest (Apperson, Laws, & Scepansky, 2006; Bockoven, 2004; DeBord, Aruguete, & Muhlig, 2004; Smith & Woody, 2000), although not necessarily student achievement (Susskind, 2005). Although presentation graphics programs may enhance the classroom experience for students, these tools also provide tremendous flexibility to instructors who no longer need to rely on preprinted materials that are neither animated nor easy to alter, functions that modern presentation software achieve easily (Christie & Collyer, 2005; Sea-
Does an Interactive WebCT Site Help Students Learn?

Joelle D. Elicker, Alison L. O’Malley, and Christine M. Williams

The University of Akron

We examined whether students with access to a supplemental course Web site enhanced with e-mail, discussion boards, and chat room capability reacted to it more positively than students who used a Web site with the same content but no communication features. Students used the Web sites on a voluntary basis. At the end of the semester, students using the enhanced site earned more points in the class than students using the basic Web site. Additionally, students using the enhanced site reported using it more often and reported higher satisfaction with the Web site, course, and instructor. We discuss practical implications of these findings.

Course Web sites supplement face-to-face classroom instruction by providing students with access to course materials, interactive textbook modules, study guides, discussion boards, journal articles, and a wealth of other resources. Course Web sites can also facilitate communication between and among instructors and students by providing their contact information and sometimes providing chat rooms or messaging services as part of their design. Internet communication generates opportunities for contact through virtual office hours, problem solving outside of class time, and the carryover of in-class discussions (Varnhagen, Drake, & Finley, 1997). This additional communication and support for learning may be especially important in introduction to psychology, as it is a gateway course to the psychology major and is often composed of first-year students who are unfamiliar with the college environment.

Research assessing students’ attitudes toward course Web sites has shown favorable results. Plous (2000) established that 85% of surveyed students reported they enjoyed having a course Web site in their psychology class and less than 1% indicated that they disliked having a course Web site (the remainder were indifferent). Despite the popularity of course Web sites and the enthusiastic response they tend to receive from students, little research has explicitly addressed their efficacy. A study by DeBord, Aruguete, and Muhlig (2004, Study 2) is a notable exception. In their examination of the effectiveness of computer-assisted teaching methods, DeBord et al. found that students in an introductory psychology course who used an optional course Web site performed similarly to those who did not use the Web site. Although they examined course Web site efficacy in terms of student use versus nonuse, DeBord et al. did not explore the specific Web site features that might be linked to enhanced course performance. We probed this issue by identifying aspects of course Web sites that encourage student use. We compared two types of course Web sites: a course WebCT site with e-mail, discussion boards, and chat room capability versus a basic course Web site that lacked these communication features. We assessed whether students with access to the course WebCT site had higher achievement, higher usage rates, and more favorable attitudes and reactions to aspects of the course compared to students with access to the basic course Web site.

WebCT offers a more comprehensive and user-friendly communication interface than a basic course Web site; users can e-mail their instructors and classmates with one mouse click. Basic course Web site users, on the other hand, must access an external e-mail system to e-mail their instructor. Furthermore, the basic course Web site did not list classmates’ contact information. We expected that students with access to the WebCT site would both use the WebCT site more often and find it easier to use than the basic course Web site because student–instructor and student–student interaction may be more easily achieved with the WebCT site. We also anticipated that students would experience more favorable reactions to WebCT than the basic course Web
site. Eighmey (1997) found Web sites perceived as more usable were also perceived as more appealing. Additionally, we explored whether these favorable reactions extended to the instructor and the course.

Use is a clear requisite for Web site efficacy, and as such we were especially interested in the implications of course Web site use for student learning. The higher rate of student usage anticipated with the WebCT site enables two important precursors for student learning: students’ increased use of pedagogical features and increased interaction with the instructor. Increased use suggests greater exposure to the Web site offerings such as test topics and interactive study modules made available for the express purpose of enhancing course performance. Additionally, easier accessibility of the instructor via the WebCT site may encourage students to resolve questions about class content and procedures outside of class, easing their concerns and increasing their understanding. We predicted that students who had access to the WebCT site would have higher achievement in the course.

Method

Participants

Introduction to psychology students (105 men and 153 women) at a large public Midwestern university participated in our study. Of the students who provided racial demographics (N = 256), 66.3% identified themselves as Caucasian, 14.7% as African American, 1.3% as Hispanic, 0.7% as Asian American, 0.7% as Native American, and 1.7% indicated “other.” First-year undergraduates comprised the majority of the sample (65.6%); 25.4% were sophomores, 6.6% were juniors, and 2.4% were seniors. Four course sections (n = 114) used the basic course Web site and four course sections (n = 144) used a section-specific WebCT site. Three instructors taught the basic course Web site sections, and two instructors taught the WebCT sections. We attempted to include in the study classes with instructors who were equivalent in terms of teaching ability. Analysis of instructor evaluations obtained prior to our data collection revealed that WebCT instructors had an average composite rating (17 items on a scale ranging from 1 to 5 with 5 indicating higher ratings) of 4.23, whereas basic course Web site instructors had an average composite rating of 4.33; a t test revealed no significant difference between these means, t(3) = 1.42, p > .05. We did not fully balance instructors across Web sites due to last-minute scheduling issues and our concern that having instructors teach two sections each (one with each type of Web site) could burden instructors and signal to students that there were different types of Web sites, possibly biasing students’ reactions to their course Web site. Students enrolled in sections without knowledge of the particular type of course Web site they would have available.

Between-group t tests revealed that students in the basic course Web site sections did not differ significantly from students in the WebCT sections in terms of their ACT scores, t(198) = −1.43, p > .05; grade point average (GPA), t(254) = −.17, p > .05; number of days absent, t(253) = .81, p > .05; or age, t(254) = 1.38, p > .05.

Procedure

Type of course Web site (i.e., WebCT vs. basic) was the fundamental basis for comparison in our study. In all classes, the syllabus described the Web site and instructors mentioned the Web site at the beginning of the semester and on occasion thereafter. Qualitative analysis of follow-up surveys assessing instructors’ behavior with regard to in-class training sessions or student reminders to use the Web site revealed no differences in instructor behavior. Student usage of the Web sites was completely voluntary. Both the basic course Web site and WebCT sites contained the following features: test topics, the course syllabus and calendar, homework assignments, paper assignment, links to psychology-related Web sites, computer-based testing information, and interactive modules with practice quizzes and tutorials from the textbook publisher. We organized these features in the same manner on both Web sites. WebCT sites had the following additional features: (a) students could directly e-mail their instructor and classmates from the WebCT site, and (b) students could interact via discussion boards and chat rooms.

Students completed paper-and-pencil questionnaires 2 weeks into the semester (N = 258), and a subset of these students (n = 155) volunteered to complete questionnaires at the end of the semester. Students indicated their reactions to aspects of the course and reported Web site usage during the second wave of data collection. We measured computer efficacy at both time periods.

Measures

Achievement. We measured students’ course achievement in terms of how many points they
received out of a possible 800. Students accrued points from their performance on five homework assignments, five exams, and one paper.

Use. Students reported whether they used the course Web site zero or one time during the semester, a few times, almost weekly, or almost daily. We used a 13-item scale adapted from Forsyth and Archer (1997) to measure the portion of time on the Web site devoted to specific tasks (e.g., communicating with the instructor, studying for tests). Participants responded on a 5-point scale ranging from 1 (none of the time) to 5 (all of the time). We used a 3-item scale developed for this study to measure students' perceptions of how easy the course Web site was to learn and use. A sample item was “The features of the course Web site were easy for me to learn.” Participants responded on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Attitudes. We adapted a 13-item scale from Forsyth and Archer (1997) to measure students’ reactions to the Web site. Sample items included “I gained valuable information from the course Web site” and “Using the course Web site was a positive learning experience.” We developed a 5-item scale to measure students’ satisfaction with the course, their instructor, and the communication they had with their instructor and classmates. Participants responded to both of these attitude measures on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Control variables. In light of findings that students’ self-reported experience with computers is related to their attitudes toward computers (e.g., DeBord et al., 2004), we anticipated that participants’ Web site usage and attitudes may be related to their computer self-efficacy. Thus, we measured computer self-efficacy for use as a potential control variable with the Computer Self-Efficacy Scale (CSES; Torkzadeh & Koufteros, 1993). The CSES consists of 30 items measured on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). Similarly, given that attendance is often highly correlated with students’ course performance (e.g., Williams & Worth, 2002; Worthington, Welsh, Archer, Mindes, & Forsyth, 1996), we expected that participants’ class attendance might be related to their achievement and therefore was appropriate as a control variable in our achievement analysis.

Results

As expected, both control variables correlated with our dependent measures. Specifically, attendance correlated with achievement, $r(255) = -.75, p < .001$, and computer self-efficacy correlated with the set of Web site use and attitude measures (correlations ranged from .18–.46, $p < .01$, $N$ ranged from 146–155). To ensure that unmodeled individual differences did not interfere with detecting the hypothesized relationships, we included attendance and computer self-efficacy as covariates in the relevant analyses.

Achievement

An ANCOVA revealed that WebCT users and basic course Web site users differed in the number of total points earned in the course when controlling for the number of days students were absent, $F(1, 251) = 9.80, p < .05, \eta^2 = .04$. WebCT students earned a mean score of 582 points ($SE = 11.36$), whereas basic Web site users earned a mean score of 529 points ($SE = 12.74$). Thus, on average the WebCT students earned 73% of the total class points, compared to 66% for the basic Web site students. This difference corresponds to an $f$ of .20, which is close to Cohen’s (1988) cutoff of .25 for a medium effect size. We conducted a post-hoc ANCOVA to assess whether increased communication with the instructor operated as a mediator and therefore functioned as an explanatory mechanism underlying this difference in achievement between the two groups. This test revealed that the significant achievement difference between the course Web site types disappeared when controlling for students’ perceptions that the course Web site improved communication with their instructor, $F(1, 149) = .50, p > .05, \eta^2 = .003$. In conjunction with our finding (reported subsequently) that Web site type was significantly related to improved communication, it appears that communication operated as a mediator of the Web site type–achievement relationship.

Use

An ANCOVA indicated that when controlling for computer efficacy, WebCT students used their course Web site significantly more than basic course Web site students, $F(1, 144) = 16.98, p < .001, \eta^2 = .11$. The WebCT mean of 2.53 corresponded to an average usage rate halfway between using the course Web site a few
times during the semester and using it on a near-weekly basis. The basic Web site mean of 1.99 indicated that on average students used the basic Web site just a few times during the semester. This difference represents a medium effect size \((f = .35;\) Cohen, 1988).

A MANCOVA controlling for computer efficacy indicated that specific usage patterns were related to type of course Web site, \(F(12, 139) = 2.89, p < .001, \eta^2 = .20\). Specifically, students using the two Web sites differed in the portion of time they spent communicating with the instructor by e-mail, \(F(1, 150) = 4.43, p < .05, \eta^2 = .03 (f = .17)\); using the textbook publisher’s Web site to do homework, \(F(1, 150) = 7.73, p < .01, \eta^2 = .06 (f = .25)\); using the textbook publisher’s Web site to study for tests, \(F(1, 150) = 7.90, p < .01, \eta^2 = .05 (f = .23)\); and checking the syllabus, reading assignments, or course calendar, \(F(1, 150) = 6.31, p < .05, \eta^2 = .05 (f = .23)\). Whereas WebCT users spent more of their time on the course Web site e-mailing their instructor and reviewing course materials, users of the basic course Web site spent a greater portion of their time utilizing the textbook publisher’s modules to do homework and prepare for tests (see Table 1). These effect sizes ranged from small to medium.

An ANCOVA controlling for computer efficacy revealed that WebCT users \((M = 4.08, SE = .09)\) more strongly endorsed perceptions that the course Web site was easy to use than did the basic course Web site users \((M = 3.82, SE = .09)\), \(F(1, 152) = 4.09, p < .05, \eta^2 = .03 (f = .17)\). This effect size was small.

### Table 1. Means and Response Frequencies Regarding Use of Web Site Features

<table>
<thead>
<tr>
<th>What Portion of Your Time on the Web Site Was Spent...</th>
<th>Type of Web Site</th>
<th>Percentage and Number of Participants (N) Indicating Each Response Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>None (1)</td>
</tr>
<tr>
<td>Communicating with your professor by e-mail</td>
<td>WebCT</td>
<td>2.02</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>1.68</td>
</tr>
<tr>
<td>Using the modules to do homework</td>
<td>WebCT</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>2.83</td>
</tr>
<tr>
<td>Using the modules to prepare for tests</td>
<td>WebCT</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>2.46</td>
</tr>
<tr>
<td>Checking the syllabus, reading assignments or course calendar</td>
<td>WebCT</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>2.60</td>
</tr>
<tr>
<td>Using the chat room or discussion boards</td>
<td>WebCT</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Note. Participants indicated their usage of various Web site features on a scale ranging from 1 (none) to 5 (all the time). All mean comparisons differed significantly at \(p < .05\). The question asking about chat room and discussion boards was not included on the basic Web site survey because these students did not have these features available.

### Attitudes

We conducted a MANCOVA controlling for computer efficacy to assess whether students’ reactions to the course Web site differed based on whether they had access to the basic or enhanced Web site. The MANCOVA revealed mean differences between course Web site types on the composite of dependent variables, \(F(13, 139) = 2.64, p < .01, \eta^2 = .20\). WebCT users and basic course Web site users differed in terms of how much they believed their Web site improved communication between themselves and their instructor, \(F(1, 151) = 23.31, p < .001, \eta^2 = .15\). The effect size for this difference was large \((f = .42)\). Students using WebCT \((M = 3.31)\) reported higher agreement with the statement, “The Web site improved communication between me and my instructor,” compared to users of the basic course Web site \((M = 2.51)\).

A MANCOVA conducted to examine whether WebCT users reported more satisfaction with the course, their instructor, the course Web site, and the communication they had with their instructor and classmates revealed that when controlling for computer efficacy, the composite of dependent variables was significantly related to the type of course Web site, \(F(5, 146) = 2.98, p < .05, \eta^2 = .09\). Compared to basic Web site users, WebCT users reported significantly higher levels of satisfaction with the course, \(F(1, 150) = 8.56, p < .01, \eta^2 = .07 (f = .27)\); instructor, \(F(1, 150) = 14.29, p < .001, \eta^2 = .09 (f = .31)\); course Web site, \(F(1, 150) = 3.50, p < .05, \eta^2 = .08 (f = .20)\).
Table 2. Mean Satisfaction Levels

<table>
<thead>
<tr>
<th>Satisfaction With</th>
<th>Type of Web Site</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>WebCT</td>
<td>4.26</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>3.85</td>
</tr>
<tr>
<td>Instructor</td>
<td>WebCT</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>4.05</td>
</tr>
<tr>
<td>Web site</td>
<td>WebCT</td>
<td>3.99</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>3.68</td>
</tr>
<tr>
<td>Communication with instructor</td>
<td>WebCT</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>3.89</td>
</tr>
</tbody>
</table>

Note. Participants responded to these items on a 5-point Likert-type scale, with higher scores indicating greater satisfaction. All mean comparisons differed significantly at \( p < .05 \). \( \eta^2 = .08 \) (\( f = .29 \)); and communication with the course instructor, \( F(1, 150) = 8.63, p < .01, \eta^2 = .07 \) (\( f = .27 \)). All of these effect sizes were medium. We present the mean levels of satisfaction for each group in Table 2.

Discussion

Our study examined whether students with access to the course WebCT site demonstrated greater learning, higher usage rates, and more positive reactions to the Web site than students with access to the basic course Web site. In line with our predictions, WebCT users earned more points in the course, found the Web site easier to use, and used the Web site more than basic Web site users. The ease with which students could contact their instructor via the WebCT site was apparently beneficial, as WebCT users indicated feeling more satisfied with communication with their instructor than basic Web site users. We also found relationships between the type of course Web site and more global satisfaction ratings; WebCT users reported greater satisfaction with their instructor and the course overall. Although the percentage of variance accounted for in each significant relationship may appear modest, the majority represented medium effect sizes (Cohen, 1988). Thus, we believe they are meaningful effects that demonstrate how course Web sites can contribute to students’ course achievement and positive course reactions. It is important to note that WebCT was not new to the campus, so instructor or student enthusiasm for a new tool is an unlikely explanation for students’ positive reactions.

WebCT students also differed from basic Web site students in terms of how they used the course Web site. WebCT users devoted more of their time to communicating with their instructor and reviewing course materials (e.g., syllabus, calendar). Users of the basic course Web site, on the other hand, primarily used it to access modules necessary for the completion of homework assignments and to obtain test preparation materials. Considered in conjunction with the higher achievement of the WebCT sections, the additional interaction that WebCT users had with their instructors may have allowed them to clarify questions about class content and procedures, continue in-class discussions, and problem solve outside of class. Our data suggest that the communication features of WebCT may be the mediating mechanism through which Web site type is related to achievement.

Further examination of Web site usage patterns rules out an alternative explanation for the observed difference in achievement. Students reported using the homework and exam review features of the basic course Web site more often than the same components in WebCT. Hence, the observed effect in performance is apparently not related to the pedagogy supplied by the textbook publishers, because the basic Web site students reported using these more but had lower performance.

A potential limitation of our study is the lack of random assignment to course Web site conditions. However, students could not self-select into sections using a specific Web site type, as this information was not available until after the class started. Additionally, we relied on self-reports of Web site usage; tracking student use on the basic Web site was not possible in light of its unrestricted and unidentifiable user access. Future research should employ objective measures of actual use and measure the specific nature of students’ online interactions with their instructor.

Our findings point to a number of aspects that instructors using course Web sites should bear in mind. First, instructors should attend to students’ computer efficacy; we found that students with higher computer efficacy had more positive reactions to the Web sites. Hence, students’ confidence in their ability to use computers may influence their reactions to particular computer programs. Instructors could measure computer efficacy at the beginning of a course and provide training as necessary. Instructors must also make their expectations for usage clear, particularly when implementing features such as discussion boards (Carswell, Thomas, Petre, Price, & Richards, 2000). Otherwise, students may experience confusion and frustration about how often and in what manner they should participate in online activities. In our study, instructors did not...
require use of these features, and the majority of students did not use the chat room or discussion boards at all. Finally, students must not perceive a supplemental course Web site as a substitute for in-class time. We found, as have others (e.g., Williams & Worth, 2002), a large negative relationship between attendance and student performance.

Our study adds to the empirical literature on course Web sites by exploring a specific feature that may make course Web sites successful: easy-to-use communication tools. Metzner, Rajeccki, and Lauer (1994) found introductory psychology coursework is vital in students’ decision to major in psychology. Our findings demonstrate that students in introduction to psychology may benefit from a course Web site with an easy-to-use student–instructor communication portal. Incorporating design features that enable easy communication into a course Web site can encourage student use and favorably impact students’ perceptions and learning. We hope our investigation helps inform the work of other instructors as they design their course sites.

References


Notes

1. We thank Rosalie J. Hall and Andrea F. Snell for their advice regarding our analyses.

2. Send correspondence to Joelle D. Elicker, Department of Psychology, The University of Akron, Akron, OH 44325-4301; e-mail: joelle@uakron.edu.
This study examined the retention of students who listened to podcasts of a primary source to the retention of students who read the source as text. We also assessed students’ preferences and study habits. Quiz scores revealed that the podcast group performed more poorly than did students who read the text. Although students initially preferred podcasts, their preferences changed immediately after the quiz. Podcasts might be a useful tool to supplement or enrich course-related material, but they are not as effective as text for delivering primary content.

The popularity of portable MP3 players as tools for students to conveniently listen to or view course lectures and content has been increasing in higher education. Audio-podcasting, akin to creating files to allow a student to listen to a lecture or reading on a tape-recorder, has become very popular, with several publishers offering content in this format. Although a number of papers report provocative uses and student enthusiasm for podcasted material (e.g., Campbell, 2005; Evans, 2008; Rosell-Aguilar, 2007) scant evidence exists with regard to actual, rather than perceived, learning impact.

Despite students’ beliefs that podcasts are effective learning tools, how well should instructors expect students to learn material presented only in an audio format? During the 1980s, cognitive psychologists studied participants’ recall after reading or listening to text. Across several studies, reading text led to better recall than listening to text (Dixon, Simon, Nowak, & Hultsch, 1982; Green, 1981; Hildyard & Olson, 1982). Scholars have reported contrary findings (e.g.,
Sannomiya, 1982, 1984), but other researchers have raised methodology questions regarding these equivocal results (see Rickheit, Strohner, Müsseler, & Nattkemper, 1987). This body of basic research raises applied questions about the use of podcasts to present primary course content, but other advantages, such as convenience, accessibility, and enjoyment, might offset these concerns.

Previous research argued that students are not very good judges of their own learning (see Dunning, Johnson, Ehrlinger, & Kruger, 2003, for a review). In a study of students’ perceived learning, use of pedagogical aids, and actual performance, Gurung and Daniel (2005) summarized reports of negligible to negative correlations between student use of such aids and student exam scores, despite positive student perceptions of the learning impact of these tools (see also Gurung, 2003, 2004). Clearly, initial student preference and self-report of learning are not the best indicators of student learning.

Students like the idea of podcasts, but do they learn primary content as well from listening to it as they do from reading it? This issue becomes more important as higher education begins to explore the possibility of audio text supplements to deliver a course’s primary content and vocabulary. This study investigated student preference and performance on podcasted versus text-based primary content to begin to provide instructors as well as publishers with guidelines and challenges for audio-podcast use. Additionally, the students who heard podcasts participated in a focus group in which they provided feedback about their learning experiences with the podcasts.

Method

Participants

Participants were 48 students (12 men, 36 women) in a developmental psychology course at a medium-sized regional university who participated as part of a course requirement. We treated all participants in accordance with American Psychological Association (APA) ethical guidelines (APA, 2002).

Materials and Procedure

We randomly assigned students to either read the 3,330-word article or listen to a 21 min, 42 sec podcast of “Mindful of Symbols” by DeLoache (2005) in preparation for a quiz. After 2 days of time to read or listen to the article, all students used a 9-point Likert scale (1 = not at all, 9 = extremely) to complete prequiz measures of their perceived knowledge and understanding of the material for the quiz, the difficulty of the material, how much they learned, and how much they enjoyed the reading or podcast. Participants also reported the amount of time they spent studying, their activities concurrent with studying (e.g., walking while listening to the podcast), the location of studying, and competing activities they performed (e.g., talking on the phone). Finally, they used the 9-point Likert scale to report the degree to which they would prefer a podcast over reading to learn important material.

All participants then completed a 10-question multiple-choice quiz about the article (e.g., “According to the author, what is the first type of symbolism that infants and young children master?”). After completion of the quiz, participants answered the last Likert scale question again.

In addition to the quantitative data collection previously described, the 23 students who heard the podcast participated in a focus group discussion immediately following the quiz and provided feedback regarding their perceptions of positive and negative aspects of podcasts as primary learning tools.

Results and Discussion

Experimental Data

Data included quiz scores, responses on the prequiz measure, and responses to the postquiz question. The 25 participants who read the article scored higher on the quiz (M = 8.16, SD = 1.11) than the 23 participants who heard the podcast (M = 5.91, SD = 1.56), t(46) = 5.78, p < .001, d = 1.70. Despite claims from many, these results reflect the basic research from the 1980s (e.g., Dixon et al., 1982; Green, 1981; Hildyard & Olson, 1982) and suggest that podcasts do not deliver primary content as well as textbooks. Students remember primary content better when they read instead of listen to it.

To explore students’ perceptions of their learning, we used a MANOVA with condition as a between-participants independent variable and students’ self-reports of their knowledge, their comprehension, the difficulty of the material, and the amount they learned from the text or podcast as dependent variables. There was a multivariate main effect for condition, Wilks’
Participants who reported doing each of the other competing activities (e.g., talking on the phone, doing other computer activities) while they were reading or listening to the article. Students who read the article reported engaging in a mean total of 3.36 (SD = 2.78) competing activities, and students who listened to the podcast reported engaging in a mean total of 2.48 (SD = 2.21) competing activities; this difference was not significant, $t(46) = 1.20, p = .23, d = .35$. As shown in Table 2, students in the podcast condition were more likely than students in the text condition to report doing other computer activities, $\chi^2(1, N = 48) = 4.17$, $p < .05, \varphi = .30$. Although students in the text condition reported a greater total of noncomputer activities such as talking on the phone, watching television, or having people present ($M = 3.12, SD = 2.59$) than

### Table 1. Means, Standard Deviations, and Univariate Statistics for Participants’ Perceptions of Learning as a Function of Condition

<table>
<thead>
<tr>
<th>Measure</th>
<th>Text</th>
<th>Podcast</th>
<th>Univariate Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>SD</td>
<td>$M$</td>
</tr>
<tr>
<td>Knew (remembered) material</td>
<td>6.56</td>
<td>0.96</td>
<td>5.52</td>
</tr>
<tr>
<td>Understood (comprehended) material</td>
<td>7.08</td>
<td>0.91</td>
<td>6.04</td>
</tr>
<tr>
<td>Difficulty level of material</td>
<td>4.00</td>
<td>1.29</td>
<td>5.04</td>
</tr>
<tr>
<td>Learn from text/podcast</td>
<td>6.40</td>
<td>0.96</td>
<td>5.78</td>
</tr>
</tbody>
</table>

Lambda, $F(4, 43) = 3.40, p < .05, \eta^2_p = .24$. As shown in Table 1, univariate results revealed that students who read the article reported that they knew more, understood more, had less difficulty, and, marginally, learned more than did students who heard the podcast. Despite these differences, students did not report spending different amounts of time reading the text ($M = 25 \text{ min}, 2.40 \text{ sec, } SD = 13 \text{ min}, 15.70 \text{ sec}) and listening to the podcast ($M = 24 \text{ min}, 20.87 \text{ sec, } SD = 7 \text{ min}, 4.88 \text{ sec}), t(46) = .22, p = .83, d = .06$, and students similarly enjoyed reading the text ($M = 5.60, SD = 1.47$) and listening to the podcast ($M = 5.48, SD = 2.04$). The difference in enjoyment was not significant, $t(46) = .24, p = .81, d = .07$. Despite claims of greater flexibility for podcasts and student preference for them when asked, students did not spend different amounts of time interacting with podcasts and text, and they did not rate them as differentially enjoyable when asked after actually interacting with the material in both media. Even with these similarities in time and enjoyment, the quiz performance difference remained.

Students reported other activities they did while they were reading and the locations for reading. The choices of activities included walking, sitting, working out, driving, doing chores, and other. Although students did not differ across all categories of reported activities for reading or listening locations as a function of condition, $\chi^2(4, N = 48) = 6.38, \text{ ns}$, 88% of participants who read the text reported sitting, and 60.9% participants who listened to the podcast reported sitting, $\chi^2(1, N = 48) = 4.70, p < .05, \varphi = .31$. Students listening to podcasts were less likely to sit and study than were students who read the material. There were no significant differences in the locations in which students read or listened to the material. Students who listened to podcasts were not more likely to take advantage of the potential for flexibility for study locations and activities provided by podcasts.

Table 2 shows the number and percentage of participants who reported doing each of the other competing activities (e.g., talking on the phone, doing other computer activities) while they were reading or listening to the article. Students who read the article reported engaging in a mean total of 3.36 (SD = 2.78) competing activities, and students who listened to the podcast reported engaging in a mean total of 2.48 (SD = 2.21) competing activities; this difference was not significant, $t(46) = 1.20, p = .23, d = .35$. As shown in Table 2, students in the podcast condition were more likely than students in the text condition to report doing other computer activities, $\chi^2(1, N = 48) = 4.17$, $p < .05, \varphi = .30$. Although students in the text condition reported a greater total of noncomputer activities such as talking on the phone, watching television, or having people present ($M = 3.12, SD = 2.59$) than

### Table 2. Number and Percentage of Participants Who Self-Reported Competing Activities as a Function of Condition

<table>
<thead>
<tr>
<th>Activity</th>
<th>Text</th>
<th>Podcast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Television</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Music</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Roommates/friends present</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Unknown people present</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Both friends and unknown people present</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Respond to instant messaging/e-mail via the Internet</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Facebook/MySpace</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Other computer activities</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Text message</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Answer phone</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Talk on phone</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td>61</td>
</tr>
</tbody>
</table>

Note. “Other” responses included “get ready,” “distracted,” “interrupted,” “other reading,” “sat with significant other,” “cleaned room,” and “showered.”

*p < .05, \varphi = .30.*
students in the podcast condition reported \((M = 2.04, \ SD = 1.89)\), this difference was not significant, \(t(46) = 1.63, p = .11, d = .47\). Students who listen to podcasts on computers do more computer activities, perhaps due to the distractions inherent in the wide variety of easily accessible computer activities. Any technological device that plays podcasts (e.g., computers, iPods, or cellular phones) might have other features that could be more interesting and distracting than the class material in the podcast.

Both before and after the quiz, participants reported the degree to which they would prefer a podcast over reading to learn important material. To evaluate responses, we used a repeated measures ANOVA with condition as a between-participants independent variable and the timing of the quiz (pre and post) as a within-participants variable. There was a main effect for condition; students preferred text \((M = 6.02, \ SD = 1.25)\) over podcasts \((M = 4.41, \ SD = 1.88)\), \(F(1, 46) = 12.31, p < .01, \eta^2_p = .21\). There was also a main effect for time; scores decreased between the pretest \((M = 5.73, \ SD = 2.09)\) and the posttest \((M = 4.77, \ SD = 1.87)\), Wilks’s Lambda, \(F(1, 46) = 19.98, p < .001, \eta^2_p = .30\). More importantly, however, a significant interaction existed between condition and time, Wilks’s Lambda, \(F(1, 46) = 18.41, p < .001, \eta^2_p = .29\). We performed simple contrasts to investigate the interaction. The difference between the pretest scores \((M = 6.04, \ SD = 1.46)\) and posttest scores \((M = 6.00, \ SD = 1.38)\) for the text group was not significant, \(t(46) = .15, p = .88, d = .04\). The difference between pretest \((M = 5.39, \ SD = 2.61)\) and posttest \((M = 3.43, \ SD = 1.34)\) for the podcast group was significant, \(t(46) = 5.38, p < .001, d = 1.55\). Although students who listened to podcasts preferred podcasts before the quiz, after the quiz the preference for podcasts decreased. Although students did not immediately learn about their performance on the quiz (i.e., their grades), merely taking the test alerted them to the limits in their comprehension after listening, and the significant change in the preferences for the podcast group reflected this realization.

**Student Focus Group Outcomes**

In a focus group following the quiz, we asked the 23 students in the podcast group how the podcasts could be made more valuable. Several issues emerged that might be helpful for future investigation as well as podcast development. Students reached a near unanimous consensus on five points: (a) The lack of signaling devices (e.g., bold words, italics) in the podcasts made it difficult to prioritize the reading and focus on the important points; (2) podcasts lack visuals such as charts and graphs that reinforce the reading; (3) the students were much less likely to review sections of the podcast than they would have been when reading it; (4) the more the voiceover in the podcast sounded like a professional reader (e.g., not casual and conversational), the less enjoyable the podcast; and (5) it was easier to listen to the podcast on the computer than to go through the trouble of downloading it to an MP3 player. Although the last suggestion might explain the results that the podcast and text groups did not differ much in where they interacted with the material, the suggestion that the podcasts might be more effective if the learners were also supplied visual support (e.g., signaled text and supporting graphs) was not tested in this study.

**Conclusions**

The results argue for caution when relying on audio podcasts to deliver primary course content. Students in the podcast group performed relatively poorly on the quiz and reported that they knew less, understood less, experienced more difficulty with the material, and, marginally, learned less than did students in the text condition. Despite the popular claims that podcasts allow for more flexibility of use, efficiency, and enjoyment (e.g., Campbell, 2005; Duke University, Office of Information Technology, 2005) and despite a lower likelihood of sitting to study, the students in this sample did not differ in where they interacted with the material, how long they studied, or how much they enjoyed the content.

As expected, students who listened to podcasts initially preferred podcasts as learning tools in this study. This finding joins an ever growing list of student preferences for pedagogy and techniques that do not positively affect their actual, as opposed to perceived, learning (e.g., Gurung & Daniel, 2005; Wesp & Miele, 2008). Student perception of learning is seldom a reliable basis for performance-based measures (e.g., Dunning et al., 2003). Interestingly, directly after taking the quiz, even without formal feedback regarding their performance, students in the podcast group realized that the podcasts were not effective tools for their learning and performance. It is possible that students are more likely to gauge the effectiveness of very poor strategies if they have experience using them and reflecting on them in an evaluated context.
The findings reported here suggest that audio podcasts are not effective learning tools for the mastery of primary course content, such as vocabulary and core concepts. The use of audio podcasts remains untested for delivering secondary content that reinforces, extends, and contextualizes the primary concepts of a course or concept. Indeed, enriching primary content in this manner might be the ideal use for audio podcasts.

References


Notes

1. We would like to thank Worth Publishers for providing the podcasts and article used in this study.
2. Send correspondence to William Douglas Woody, School of Psychological Sciences, University of Northern Colorado, Greeley, CO 80639; e-mail: william.woody@unco.edu
the candidates’ students and letter writers and live public speaking performances. In this case, excellence in teaching is not defined a priori; instead, it is construed as the search process unfolds.

References


Notes

1. We thank Jason Sikorski for helpful comments and suggestions on an earlier draft of the article and Scott Wheeler for assisting with data collection. We also are grateful to Randolph A. Smith and three anonymous reviewers for their helpful comments on earlier versions of this article.

2. Send correspondence to William Buskist, Psychology Department, Auburn University, Auburn, AL 36830–5214; e-mail: buskisw@auburn.edu.

Providing Students With Instructors’ Notes: Problems With Reading, Studying, and Attendance

Michael A. Vandehey, Crystale M. Marsh, and George M. Diekhoff
Midwestern State University

Researchers have found that instructor-provided notes improved student performance but were criticized for a lack of ecological validity in the research designs. This study improved on past research by using instructor-provided notes in a required class throughout an entire semester. Previous findings were not supported in that final grades and attendance were similar across student-generated notes, instructor-provided partial notes, and instructor-provided full notes. Students evidenced poor reading, studying, and attendance behaviors in all 3 conditions.

Anderson and Armbruster (1991) reported that students generally record only 50% to 70% of the important ideas from a lecture. Some teachers use lecture handouts to aid students (Hartley, 1976), and improvements in recall were evident with either full or partial instructor-provided notes (Kiewra, Mayer, Christensen, Kim, & Risch, 1991).

Ideally, students would read instructor-provided notes prior to the lecture so that they could be better prepared to learn, ask well-formed questions, and answer questions (Charman & Fullerton, 1995). Concerns about providing students with notes included the findings that only 31% to 40% of students reviewed these notes before the lecture (Charman & Fullerton, 1995; Murphy & Cross, 2002) and absenteeism (Potts, 1993; Russell, Caris, Harris, & Hendricson, 1983). On a positive note, Herson, Sosabowski, and Lloyd (1999) found that only 13% of pharmacy students reported missing one or two classes due to the availability of notes.

Researchers found instructor-provided partial notes to be the best solution to concerns over absenteeism and inattention. Partial notes guide students to important concepts, facilitate learning by requiring students to fill in extra information, and engage students in thinking about the material (Hartley, 1976; Katayama & Robinson, 2000; Kiewra et al., 1991).

Limitations of Previous Research

Previous research identified important variables affecting student learning with instructor-provided notes. However, most of the studies, in applying scientific controls, gained internal validity by sacrificing external validity to the classroom and typical student behavior (Bligh, 1998; Collingwood & Hughes, 1978; Hartley, 1976; Locke, 1977). Hartley (1998) wrote, “These early studies suffered from a lack of ‘ecological validity’—that is to say, they were experimental in nature, and not very realistic in practice” (p. 81). Collingwood and Hughes (1978) listed the following external validity violations of lecture-note studies: (a) notes studies are artificial and unlike actual class situations, (b) students participate to meet a course...
requirement or earn extra credit, (c) students receive information via an audiotape or videotape recording, (d) students listen to an artificially short lecture session(s), (e) students take posttests at unusual times (e.g., immediately after or within a few days of the lecture), (f) students turn in their notes to prevent revision or review, and (g) the subject matter is not always related to course material (see Hartley, 1976; Hohn, Gallagher, & Byrne, 1990; Katayama & Robinson, 2000; Kiewra, 1985; Kiewra, DuBois, Christian, & McShane, 1988; Kiewra & Frank, 1988; Kiewra et al., 1997; Mayer, 1983; Maqsud, 1980; Thomas, 1978). Finally, we would add that some studies allowed review time before the posttest (see Hohn et al., 1990; Kiewra et al., 1988; Kiewra & Frank, 1988; Kiewra et al., 1997; Maqsud, 1980; Ruffini, 1999).

Rationale and Purpose of This Study

Our study improved on previous research by addressing the external validity concerns of Collingwood and Hughes (1978) and Hartley (1998). Data collection lasted for an entire semester and the material was class relevant and presented during normal class periods. Also, students kept their notes and studied them according to personal schedules.

We randomly assigned classes to one of three conditions: student-generated notes (SGN), instructor-provided partial notes (IPN), and instructor-provided full notes (IFN). We hypothesized that students in the IPN group would have the highest grades (Collingwood & Hughes, 1978; Katayama & Robinson, 2000; Kiewra et al., 1988; Kiewra et al., 1991; Russell et al., 1983) and that students in the IFN group would have the lowest attendance rates (Potts, 1993; Russell et al., 1983).

Method

Sample

Two-hundred forty-five undergraduates met the criteria for inclusion in the study. Participants had to be 18 years of age or older and enrolled in general psychology, and they volunteered for the study. Students in classes with provided notes used passwords to access the Internet site regardless of their participation in the study.

The sample consisted of 61% women and 39% men. Ages ranged from 18 to 51, with 61% being 18 or 19, 28% being 20 to 24, and 12% being older than 25. The majority of participants were White (76%), followed by African American (8%), Hispanic (7%), Asian (4%), Caribbean (3%), and other (3%). A cross-section of students is evidenced by the listing of 54 majors; 7% of students listed psychology as their major.

Implementation

Five instructors taught from the same lecture notes, overheads, similar in-class exercises, and used the same tests. Eleven classes participated in one of three conditions: (a) SGN: 4 classes, (b) IPN: 4 classes, and (c) IFN: 3 classes. Each instructor taught a note-type condition only once. Instructors met biweekly to discuss lectures and in-class exercises.

Participants filled out a demographic survey, signed attendance sheets, completed five exams, and filled out an exit survey. Attendance was voluntary and not a part of the class grade. The final grade consisted of totaling the raw scores from each exam. Each test contained 50 multiple-choice questions covering two chapters.

Results

The effect of one independent variable (SGN, IPN, and IFN) on final grade and attendance was of interest in this research. We used two between-subject one-way ANOVAs to evaluate the data. Means and standard deviations for final grade and attendance appear in Table 1. Type of notes had no significant effect on either final grades, F(2, 242) = .23, p = .80, or attendance, F(2, 242) = .61, p = .55. More powerful a priori comparisons also failed to identify any differences as statistically significant.

Sixty-four students failed to turn in the exit interview. We used two-tailed independent-samples t tests to assess whether a difference in final grade existed between those students who did (M = 174.92, SD = 25.9) and did not (M = 176.75, SD = 25.99) complete the exit interview, t(243) = 0.49, p = .63. A second two-tailed independent-samples t test assessed whether a difference in attendance existed between those students who did (M = 0.72, SD = 0.22) and did not (M = 0.67, SD = 0.23) complete the exit interview, t(243) = −1.38, p = .17. There was no difference between the two groups on final grades or attendance.

The following self-report data are from the exit survey. One-hundred thirty-four (55%) students correctly completed, 64 (26%) students failed to complete, and 47 (19%) students failed to complete parts of the exit survey. Therefore, sample size varies from question to question. Self-reported reviewing, reading, and studying behaviors (Table 2) offer insight into the lack of effect type of notes had on final grades and attendance. Despite 72% of students (n = 118) having downloaded the IPNs prior to the lectures, most students did not review them before the lectures. Additionally, students missed an average of 9 to 10 classes.

Discussion

This study sought to determine the effect that type of class notes had on final grades and semester-long attendance. In contrast to previous research, we did not find support for the hypothesis that students with IPNs would have the highest

| Table 1. Means and Standard Deviations for Type of Notes and Attendance |
|-----------------------------|-------------------|-------------|
|                             | M                | SD          | n  |
| SGN                         | Final grade       | 174.2       | 24.5 | 48 |
|                             | Attendance        | 0.70        | 0.23 |    |
| IPN                         | Final grade       | 174.8       | 26.5 | 120|
|                             | Attendance        | 0.72        | 0.21 |    |
| IFN                         | Final grade       | 177         | 26.1 | 77 |
|                             | Attendance        | 0.69        | 0.25 |    |

Note. SGN = student generated notes; IPN = instructor-provided partial notes; IFN = instructor-provided full notes.
grades (Collingwood & Hughes, 1978; Katayama & Robinson, 2000; Kiewra et al., 1988; Kiewra et al., 1991; Russell et al., 1983) nor the hypothesis that students with IFNs would have the lowest attendance (Potts, 1993; Russell et al., 1983).

Consistent with past research (Charman & Fullerton, 1995; Murphy & Cross, 2002), 72% of our students downloaded notes prior to the lecture, but 58% to 61% of the students with IPNs did not review them before hearing the corresponding lecture. Furthermore, 8% to 15% of the students did not read the text, 53% to 62% read only parts of the chapter, 23% to 40% of students reported studying only the night before the test, and 3% to 5% of students in the SGN and IFN conditions reported not studying at all. Although previous research in laboratory settings identified and isolated important factors that improve student performance, this study suggests that students are not employing the tools appropriately or in a timely manner to improve their grades.

Russell et al. (1983) and Potts (1993) were concerned that students receiving full notes might attend class less. A surprising finding of our study was that possessing IPNs did not impact attendance. Unlike Herson et al. (1999), who found good attendance of students with IPNs, attendance was poor in all three note-type conditions.

Limitations and Recommendations

Applied research in this area needs to continue; however, researchers may need to select participants carefully. For example, 61% of our entire sample was 18 or 19 years old and represented 54 different majors. The motivation to perform well as a freshman in a nonmajor course may have been low. Future researchers may want to look at upper level undergraduates in major-specific coursework.

Collingwood and Hughes (1978) considered the use of audiotaped lectures (more recent research used videotaped lectures) to be a violation of external validity despite the fact that audio or videotapes control for presentation style. We used five instructors to address this concern; however, varying teaching styles may have confounded the impact of provided notes. Future applied researchers may want to control for instructor variables.

Special populations may also be of interest. Nontraditional (age 25+), nonnative speaking, and commuting students may use IPNs differently than the typical student. A limitation of this study may be that the sample was too diverse as IPNs may be beneficial for some student populations but not other student populations.

The findings of experimental note-type studies make sense in light of our applied study. Note type may well affect grades in situations in which note use, study time, and individual differences are controlled, but these circumstances do not define the real college learning environment.

References


High School Psychology and Student Performance in the College Introductory Psychology Course

Marcia Rossi
Tuskegee University

Jared Keeley and William Buskist
Auburn University

We compared performance in the introductory psychology course at 2 universities between students who had taken a high school course in psychology relative to those who had not. Having taken a high school psychology course provided little practical benefit in terms of academic performance in the college-level course. These results run counter to Hakala and Ernst’s (2003) recent findings that the teaching of the high school course has improved considerably in the past decade. Research into this discrepancy is likely to yield further improvements in the teaching of the high school course.

Each year roughly 505,000 students take an introductory psychology course in high school (Hakala & Ernst, 2003). Although the high school course has been in existence for well over a century (Engle, 1967), research has called into question its usefulness in preparing students for the college-level introductory course (e.g., Dambrot & Popplestone, 1975). In fact, Griggs, Jackson, and Meyer (1989) proclaimed that the high school and college courses exist in “two different worlds” (p. 118).

These concerns helped propel a set of actions intended to reevaluate and enhance high school psychology courses. In 1992, the Teachers of Psychology in Secondary Schools (TOPSS) was created to reform and standardize high school psychology curricula, increase enrollment in psychology, and develop a support network for teachers (Ernst & Petrossian, 1996). TOPSS has since worked toward those goals by, for example, publishing curriculum materials, establishing essay contests for outstanding students, providing workshops for teachers, and founding the Psi Alpha honor society. Shortly thereafter, the American Psychological Association (APA, 1999) published its national high school psychology standards. Bristol and Gillis (2001) noted that such reforms are essential to upgrading the quality of high school psychology and echoed Hakala's (1999) observation that such changes will improve the high school course.

Recently, Hakala and Ernst (2003) conducted a national survey of high school psychology teachers to determine whether the qualifications and practices of these teachers have changed in the decade since TOPPS was established. Hakala and Ernst concluded that positive changes in the teaching of the high school course have occurred including, among other things, more emphasis on research methodology and biological psychology.

What Hakala and Ernst’s (2003) study did not provide, though, is any sort of assessment as to how such reported advancements in the teaching of the course influenced student learning. In light of reforms in the teaching of the high school psychology course, we sought to reexamine how students who take the high school course perform in the college course relative to their counterparts who have not taken the high school course.

Method

Participants

We recruited students from introductory psychology courses from two universities in Fall 2002. Tuskegee University is a small, private, historically Black university (n = 123) and Auburn University is a large, land-grant, predominantly White university (n = 350). We included both lower and upperclassmen in our analysis, except for those students who had taken a previous college psychology course (Tuskegee = 14, Auburn = 30). In total, only 8 students (1 from Tuskegee and 7 from Auburn) had taken the high school Advanced Placement course. We dropped these students from the analysis because such a small number of participants does not represent a sufficient data set for a separate meaningful statistical analysis. Therefore, after dropping these participants, we analyzed the data from 108 students from Tuskegee and 313 from Auburn.

Measures and Procedure

During the first week of class, participants at both institutions completed a one-page demographic survey and a 100-item multiple-choice examination, comparable to a comprehensive introductory psychology final exam. We also gathered students’ final grades after completion of the course. At Auburn University, participants also completed a 100-item final exam at the end of the course that we matched for content to the pretest. Based on their responses to one question on the demographics survey, we divided students from both institutions into two groups: students who had taken a high school course in psychology and those who had not. At Tuskegee University, 19 students had taken the high school course and 89 had not; at Auburn University, 95 stu-
FACULTY FORUM

If You Post It, Will They Come? Lecture Availability in Introductory Psychology

M. Christina Hove and Kevin J. Corcoran
University of Cincinnati

Web-enhanced educational programs such as Blackboard (2003; http://www.blackboard.com/) provide opportunities for instructors to make supplemental course materials available to students. However, little research has investigated the effects of unlimited access to course lectures on achievement and attendance in traditional postsecondary classroom settings. Thus, we investigated the effect of lecture presentation availability on class attendance and academic performance in 2 sections of introductory psychology courses. Students with unlimited access to lecture presentations earned significantly higher grades than students who did not have similar access. Although we did not find significant differences in attendance between classes, attendance moderated the relation between class format and course grade. We discuss further implications and future research.

Educators are increasingly using academic Web sites to supplement information presented in the classroom (Campus Computing Project, 2001). Research suggests that psychology faculty view the Internet as an effective teaching tool (Vodanovich & Piotrowski, 2001). Consistent with these findings, a recent national survey found that 72.1% of college seniors reported the use of presentation software in their classes, 67.4% reported the use of Web-enhanced educational programs in their classes, and 62.2% reported the use of dedicated Web sites in their classes (Katz, 2006). Although previous research has examined the educational efficacy of supplementing (Brothen & Wambach, 2001; DeBord, Aruguete, & Muhlig, 2004; Grimstad & Grabe, 2004; Heffner & Cohen, 2005) with online materials, research has not examined the effects of making lecture presentations available via the Web to students in traditional postsecondary classes. As educational technologies, such as the Internet, provide greater opportunities for unlimited access to course materials, professors are left to question whether such increased availability will improve or impede academic achievement.

Previous research regarding the effects of supplemental course materials on academic achievement has produced inconsistent results. Some studies have not found significant academic gains associated with supplementing traditional postsecondary class procedures with online materials. For example, Brothen and Wambach (2001) found a negative correlation between online study questions and quiz scores. Similarly, DeBord et al. (2004) did not find academic gains associated with optional Web site activities supplementary to a traditional college class. Conversely, other studies have demonstrated academic gains associated with the supplementation of traditional classes and class materials. Grimstad and Grabe (2004) found that college students who voluntarily completed supplemental online questions related to course content obtained significantly higher grades than students who did not use the supplemental questions. Similarly, Heffner and Cohen (2005) found positive correlations between academic performance and supplemental course-specific Web sites. As these studies demonstrate, it is difficult to come to any conclusions regarding the efficacy of supplementing traditional postsecondary classes with online materials based on the existing body of literature.
Importantly, some research suggests that providing supplemental content materials may decrease attendance among postsecondary students. Previous research has found attendance to be positively correlated with learning and academic achievement (Jones, 1984; Launius, 1997; Van Blerkom, 1992; Vidler, 1980). In a recent survey, Friedman, Rodriguez, and McComb (2001) found that among college students high in absenteeism the availability of class materials from another source (e.g., text, Web, tutor, classmate) was the third most frequently cited reason for non-illness- and non-emergency-related absences. Conversely, other research has found that students continued to perceive benefits associated with face-to-face instruction, even when course materials were available from another source, such as the Internet (Foust, Cruickshank, Stringer, & Olander, 1999). Our review of the literature did not reveal any research that investigated student absenteeism associated with the availability of supplemental online class materials, in particular with the instructor’s Microsoft PowerPoint lectures. These gaps in the literature prompted a question: If you post it, will they come?

Providing students with unlimited opportunity to access online lecture presentations from a college class has the potential to influence academic performance in a number of ways. The availability of lecture presentations may help improve academic performance by providing students with a supplemental source of information and additional study aids. Alternatively, making lecture presentations available may encourage absenteeism by providing students with a way to get lecture materials without going to class (Friedman et al., 2001). Conversely, the availability of online lecture materials may allow students to compensate for absences by providing them with access to class information they would not otherwise have, which may counter the negative association between absenteeism and academic performance (Jones, 1984; Launius, 1997; Van Blerkom, 1992; Vidler, 1980). Because little research has addressed this particular issue, we investigated the effects of unlimited access to supplemental class materials on grades and attendance in a traditional college class.

Based on previous research (Grimstad & Grabe, 2004; Heffner & Cohen, 2005), which found academic achievement gains associated with supplemental course-specific Web sites, we hypothesized that students with online access to lecture presentations would earn significantly more points on exams than students without online access to lecture presentations, producing significantly higher overall course grades. Due to the conflict in the literature regarding the effects of supplemental class materials on attendance (Foust et al., 1999; Friedman et al., 2001), we also explored whether students with online access to lecture presentations would attend class at a different rate relative to students without online access to lecture presentations.

Method

Participants and Procedure

Participants were 365 students (47.7% women, 52.3% men) enrolled in two sections of an introductory psychology course at a large Midwestern university.

We compared two sections of an introductory psychology course that met on Tuesday and Thursday afternoons during one academic quarter. We randomly assigned online availability of lecture presentations to the early afternoon class (n = 233), or the supplemented class. The midafternoon class (n = 132), or the traditional class, received standard PowerPoint-enhanced lecture presentations during its scheduled class time and was able to access review sheets via the university’s Blackboard 5 Web site (2003; http://www.blackboard.com/), which required the use of a unique student identification code. The supplemented class received the same PowerPoint-enhanced lecture presentations during class as well as being provided with unlimited access to the lecture presentations and review sheets via another Blackboard 5 Web site, which also required unique access codes. In brief, the supplemented class could view the lecture presentations anytime through the Web, whereas the traditional class could view the lecture presentations only by attending class.

We controlled for a number of variables across the two classes, including course instructor, lecture presentations, content of lectures, review sheets, and exams. We presented both classes with standard lectures supplemented with Microsoft PowerPoint presentations, which included material not covered in the textbook, during regular class meetings. The university’s institutional review board approved all aspects of our study.

Measures

Course grade. The accumulated points from four exams totaling 300 points constituted course grade. Exams were identical for the supplemented and traditional class.
**Attendance.** The instructor took attendance every third class, resulting in seven measures of attendance.

**Results**

The first hypothesis, which predicted that students with online access to lecture presentations would earn more course examination points than students without online access to lecture presentations, demonstrated by overall course grades, was supported by the data. We examined mean differences between the supplemented and traditional classes on accumulated course examination points using an independent samples *t* test. We found a significant difference between the two classes, \( t(202) = 2.08, p < .05, \) Cohen’s *d* = 0.25 (small effect; Cohen, 1992), such that students in the supplemented class (\( M \) total points = 237.54, \( SD = 30.05 \) obtained significantly higher grades than students in the traditional class (\( M \) total points = 228.60, \( SD = 43.76 \), with 300 total points possible.

Our exploration of whether students in the supplemented class would attend class at a different rate relative to students in the traditional class did not reveal significant differences in attendance rates between the two classes, \( t(363) = .59, \) ns. Students in the supplemented class were present a mean of 3.29 (\( SD = 1.62 \) days out of 7 days the instructor took attendance. Similarly, students in the traditional class were present a mean of 3.20 (\( SD = 1.43 \) days out of 7 days.

We found that attendance was positively correlated with academic achievement regardless of access to online lecture presentations. Our analyses revealed a significant correlation between attendance and course grade in the supplemented class, \( r(233) = 0.39, p < .01, \) and the traditional class, \( r(132) = .47, p < .01.\) We used hierarchical multiple regression analysis to determine if attendance moderated the relation between class format (supplemented or traditional) and course grades. Following the procedure described by Aiken and West (1991), we mean centered the attendance values. On Step 1 of the regression we entered class and attendance main effects, and on Step 2 we entered the interaction between class and attendance. We calculated effect sizes (Cohen’s *d*) using the formula \( d = 2t/\sqrt{df} \) (Rosenthal & Rosnow, 1991). The regression results appear in Table 1.

Results at Step 1 indicated that class format and attendance were both related to higher overall course grades. At Step 2, we found a significant two-way interaction between attendance and class format (\( \beta = -0.26, p < .01 \), suggesting that attendance moderated the relation between class format and course grade (Baron & Kenny, 1986).

Figure 1 presents the predicted cell means, which we derived from the regression equation, with higher and lower values represented as 1 SD above and below the respective centered means (Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003). Our findings suggest that the supplemental class format was particularly useful for students with lower attendance, whereas among students with higher attendance, class format was not influential in determining course grades.

**Discussion**

We were interested in examining whether making lecture presentations available to students would produce learning and academic gains. Our results suggest
that posting lecture presentations to supplement a traditional college class was associated with a small, positive effect on academic achievement. Our results support previous research (Grimstad & Grabe, 2004; Heffner & Cohen, 2005), which found improved academic achievement associated with supplemental course-specific Web sites.

We were also interested in investigating the effects of unrestricted access to lecture presentations on students’ attendance. We found that unlimited access to lecture presentations did not negatively affect students’ attendance rates. Consistent with previous research (Jones, 1984; Launius, 1997; Van Blerkom, 1992; Vidler, 1980), we found that higher rates of class attendance were associated with higher overall course grades. Interestingly, our findings suggest that the online availability of lecture presentations may have countered the negative association between absenteeism and academic performance. We found that the supplemented class format was associated with higher overall course grades among students with lower rates of attendance. We believe that the availability of online lecture materials may have allowed students in the supplemented course to compensate for absences by providing class information they would not otherwise have had.

We would like to acknowledge limitations of this study. Most important, our investigation employed a quasi-experimental design, and caution should be exercised in the generalization of results derived from our study. The quasi-experimental design limits our ability to assess or control for a number of factors, including variations in academic ability and motivation. Thus, we view our study as a preliminary investigation that requires additional research to more fully elucidate the causal relationships involved in this phenomenon. Future studies should assess important characteristics, such as variations in academic ability and motivation that would provide greater comparability and conclusiveness. Additionally, future studies aimed at exploring the optimal content to incorporate into online course supplementation, including links to additional information posted on external Web sites, are required to fully appreciate the academic benefits offered by this platform. The individual and dynamic nature of learning means that there will likely always be additional studies required to fully understand the learning process and how tools like the Internet impact that process.

We view this research as one of the first steps in more clearly understanding the benefits and drawbacks to technologically enhanced learning in postsecondary education. Our results suggest that providing online access to lecture presentations may be academically beneficial to college students. The availability of online lecture presentations may help to improve academic performance by providing students with a supplemental source of information and additional study aids. Such supplemental information may be particularly useful among students with lower rates of attendance, who would not have access to relevant lecture content otherwise.

References


**Notes**

1. M. Christina Hove is now at the Center for the Study of Health and Risk Behaviors at the University of Washington. Kevin J. Corcoran is now at the School of Arts and Sciences at Northern Kentucky University.
2. Manuscript preparation was supported in part by National Institute for Alcohol Abuse and Alcoholism Grant T32AA007455.
3. We thank Randolph A. Smith and three anonymous reviewers for their helpful feedback during the editorial process. We also thank Mary E. Larimer and Matthew Funke for their assistance and many contributions.
4. Send correspondence to M. Christina Hove, Center for the Study of Health and Risk Behaviors, Department of Psychiatry & Behavioral Sciences, University of Washington, 4225 Roosevelt Way NE, Box 354694, Seattle, WA 98105-6099; e-mail: MChristinaHove@gmail.com.
Differential Effects of Full and Partial Notes on Learning Outcomes and Attendance

Tara L. Cornelius and Jamie Owen-DeSchryver
Grand Valley State University

Although college instructors are increasingly providing students with online notes, research is equivocal on how such notes affect student outcomes. This study examined partial versus full notes in introductory psychology classes while controlling for initial levels of student knowledge and academic ability. Results suggested that students receiving partial notes performed better on examinations later in the semester and on conceptual questions during the cumulative final examination than students receiving full notes. Students receiving full notes also self-reported more negative effects on attendance. We provide possible interpretations of these data and suggest areas for further investigation.

As early as the 1920s, researchers examined differences among students based on their note-taking strategies and skills (Crawford, 1925). Experimental research has demonstrated that taking notes during lectures improves performance on free recall tasks (Fisher & Harris, 1973; Maqsud, 1980) as well as on typical student outcome measures, including tests and final grades (Di Vesta & Gray, 1972; Kiewra, DuBois, Christensen, Kim, & Lindberg, 1989; Kiewra et al., 1991; Van Meter, Yokoi, & Pressley, 1994). However, the mechanism through which note taking facilitates learning in undergraduate students remains unclear. One possibility is that notes serve as an external memory device (Miller, Galanter, & Pribram, 1960). From this perspective, it is not the act of taking notes, but the availability of the notes for later review that results in student learning. Another possibility is that note taking encourages encoding of lecture material (Di Vesta & Gray, 1972; Pardini, Domizi, Forbes, & Fettis, 2005). Through the process of taking notes, students are able to personalize and reorganize information, a strategy that assists in encoding and elaboration. Fisher and Harris (1973) further suggested that note taking might serve both of these functions, as these researchers demonstrated that recall was highest for students who both took and reviewed their notes.

Recent technological advances have significantly altered how students take notes in college classrooms. New technologies allow instructors to provide notes to their students via Blackboard (www.blackboard.com) or other similar Web-based systems. This practice may allow instructors to provide more response opportunities for students, because instructors can use lecture time to engage and query students, rather than waiting for students to copy notes from overheads or slides (Austin, Lee, Thibeault, Carr, & Bailey, 2002). Simultaneously, when instructors provide lecture notes, students should have fewer transcription errors and omissions, which is important given research that has shown that students may fail to record up to 50% of relevant lecture information (Anderson & Armbruster, 1991). Additionally, students report that instructor-provided notes allow them to listen more intently during lectures by decreasing time spent “copying” information (Mantei, 2000).

Although research has suggested that students perceive that instructor-provided notes are useful to their learning and that they “like” having the handouts (Frey & Birnbaum, 2002; Pardini et al., 2005), the question remains as to whether these notes actually enhance learning outcomes. In a survey study on the use of instructor-provided notes, students reported that having access to notes contributed to procrastination in learning activities and hindered note taking during lectures (Pardini et al., 2005). Additionally, providing notes to students may encourage absenteeism if students do not perceive the benefits of class attendance (Potts, 1993). Results reported by both Murphy and Cross (2002) and Weatherly, Grabe, and Arthur (2002–2003) suggested that students with access to notes earned lower final grades than students
without access; however, the designs in both studies were confounded by the fact that the students may have differed from one another at the outset. Murphy and Cross identified groups based on self-described usage of instructor-provided notes, and Weatherly et al. compared students across semesters, leaving open the possibility that students in different semesters may have had varying levels of skill or that the timing of the class may have attracted different types of students. Despite these methodological issues, the collective results of these studies do suggest that instructor-provided notes may have a detrimental impact on student learning and thus indicate a need for thoughtful, well-researched approaches on how to best provide notes to students in college classrooms.

One aspect of instructor-provided notes worthy of close consideration is the degree of completeness of such notes. Instructors who choose to provide students with notes may offer full notes, including all lecture materials, diagrams, and content-based information, or they may choose to provide partial notes, where students receive an outline, guide, or incomplete framework into which they insert information while listening to the lecture. Although empirical research on this topic is limited, one recent study did suggest that given a choice, students preferred partial over full notes (Grabe, Christopherson, & Douglas, 2004–2005). This preference for partial notes may be important, as Annis (1981) suggested that students scored better on both essay and multiple-choice exams when they accessed partial notes or recorded their own notes as compared with when they received full notes. A more recent study by Vandehey, Marsh, and Diekhoff (2005) examined the differential effectiveness of student-generated notes, instructor-provided partial notes, and instructor-provided full notes on final grades and attendance. In this study, five different instructors used identical lectures and classroom activities, and researchers evaluated which of the note-taking strategies was most beneficial. Contrary to the hypotheses of the authors, results suggested that type of notes did not have a significant effect on either final grades or attendance.

Rationale and Purpose of the This Study

Our study adds to the existing body of research assessing the impact of instructor-provided notes on learning outcomes for college students, while expanding the work of Vandehey et al. (2005) by addressing several variables that may impact final grades, including general academic performance and prior knowledge of psychological principles. Due to concerns about students' perceptions of being “deprived” of notes or contamination across sections, we chose not to include a condition in which students did not receive any instructor-provided notes; thus we had two conditions: instructor-provided full notes (IFN) and instructor-provided partial notes (IPN). We hypothesized that students receiving partial notes would demonstrate better overall performance on examinations and final grades, as well as better class attendance. We also hypothesized that students receiving partial notes would score comparably higher on conceptual and application exam questions, because they elaborated on information through the act of taking notes. Furthermore, we expanded on previous research by maintaining the variability in content and outcome measures typically demonstrated in college teaching, controlling for instructor variables through both instructors teaching both experimental conditions, statistically controlling for prior knowledge and general academic ability, and examining student performance on different types of outcome measures, including test questions that assessed definitional knowledge and those that were conceptual or applied in nature.

Method

Participants

Undergraduate students (N = 307) enrolled in four sections of introductory psychology participated in this study. Of the total sample, 153 students received IFN and 154 students received IPN. Participants were at least 18 years of age, enrolled in the authors’ sections of the course, and signed a consent document to be included in the study. The sample consisted of 219 women and 88 men, with an average age of 18.64 (range = 18-38). The majority of participants were full-time students (n = 297); employed (n = 158); worked, on average, 9.02 hr a week; and listed their academic standing as freshmen (n = 218). Approximately one third of participants identified their major as undeclared (n = 101).

Implementation

Of the four introductory psychology sections assessed, the first and second authors each taught two sections. We randomly assigned students from one section of the first author’s course to IFN, and students in
the other section received IPN. To control for possible variation in the composition of students selecting courses in the morning versus afternoon, we balanced experimental assignment across time of day for the second author’s sections; thus, students in the first author’s morning section received partial notes, and students in the second author’s morning section received full notes. Students assigned to the IFN condition received all the instructor notes, posted on Blackboard in Microsoft PowerPoint format; we instructed students to download these notes daily and bring them to all class sessions. Students assigned to the IPN condition received partial notes, consisting largely of headings and titles of definitions and concepts, which required students to add information to complete the notes. We also provided the partial notes on Blackboard in PowerPoint and instructed students to bring these notes to every class session. Class sizes and the number of class sessions were equivalent across conditions. Information acquired on the exit interview revealed that the majority of participants in both conditions (78.75% in IFN, 75.45% in IPN) downloaded and used the notes during most class sessions.

Students completed a basic demographic measure, a knowledge-based pretest of psychological concepts, and four exams. Students also participated in random attendance probes and completed an exit survey on their use of the instructor-provided notes. Additionally, we examined ACT scores to determine if there were any preexisting differences between students in the four sections on this approximation of academic aptitude.

Results

The primary dependent variables were grades on examinations, final grades, and attendance. We also evaluated exam scores, post hoc, because we speculated that performance on different types of examination questions (i.e., definitional vs. conceptual) might vary across the conditions of instructor-provided notes. Definitional questions were those questions that required rote memorization of factual or historical information. Conceptual questions were those that required application of a theoretical concept to an example that required additional mastery of the material beyond the definition.

To ensure that the two groups were roughly equivalent prior to the course, we assessed demographic characteristics and pretest performance measures. There were no statistically significant differences between the two groups on initial knowledge for psychological concepts (as measured by the pretest), age, gender, hours worked, credits, or years of college. However, there was a significant difference between groups on ACT scores, with students receiving IFN having higher average ACT scores, \( t(305) = -2.27, p = .024 \), than students assigned to IPN, which was a small to medium effect \((d = -0.26)\). Table 1 provides the means and

<table>
<thead>
<tr>
<th>Table 1. Descriptive Statistics for Dependent Variables and ACT Scores Across Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ACT*</td>
</tr>
<tr>
<td>Exam 1 (Overall)</td>
</tr>
<tr>
<td>Conceptual</td>
</tr>
<tr>
<td>Definitional</td>
</tr>
<tr>
<td>Exam 2 (Overall)</td>
</tr>
<tr>
<td>Conceptual</td>
</tr>
<tr>
<td>Definitional</td>
</tr>
<tr>
<td>Exam 3 (Overall)*</td>
</tr>
<tr>
<td>Conceptual</td>
</tr>
<tr>
<td>Definitional</td>
</tr>
<tr>
<td>Exam 4 (Overall)*</td>
</tr>
<tr>
<td>Conceptual*</td>
</tr>
<tr>
<td>Definitional*</td>
</tr>
<tr>
<td>Final grades*</td>
</tr>
<tr>
<td>Attendance</td>
</tr>
</tbody>
</table>

\(^a n = 153, ^b n = 154. \quad ^* p \leq .05.\)
standard deviations of each of the dependent variables and ACT scores of each group.

We examined the relation between ACT scores and the dependent variables using Pearson correlation coefficients. ACT scores were significantly correlated with all outcome measures (r values ranged from .26 to .31 for exams and final grades, p < .001 for all analyses). Because ACT scores were significantly correlated with all examination scores and final grades, we used ANCOVA, with ACT scores as the covariate, to analyze performance on examinations and final grades. This procedure controls for and removes variance already accounted for by the covariate, and because ANCOVA assumes homogeneity of regression, we also conducted homogeneity and linearity of slope tests. These tests revealed nonsignificant results, with interaction term partial η² ranging from .003 to .017. Using ANCOVA, we analyzed each of the four examination scores to determine differences between the two groups. These analyses revealed nonsignificant results for Exams 1 and 2, F(1, 304) = 2.73, ns; F(1, 304) = 1.51, ns, but significant differences for Exam 3 and Exam 4, F(1, 304) = 5.45, p = .02; F(1, 304) = 5.05, p = .025, with students assigned to IPN scoring better on these two exams compared to students assigned to IFN. Partial η² and Cohen’s d indicated small to medium effect sizes for both Exam 3 and Exam 4 (η² = .018, d = .20; η² = .016, d = .22, respectively).

Examination questions were classified as definitional or conceptual according to the test bank from which they were derived, or in cases where the instructor created the examination question, both the instructor and a student research assistant conferred to designate questions as conceptual or definitional. We calculated student performance on examinations for the two domains of questions and used ANCOVA to determine differences between students in the IFN group and students in the IPN group on conceptual and definitional questions. The results revealed no statistically significant differences between those domains of questions for any of the unit examinations. However, on the final examination, Exam 4, students assigned to IPN scored significantly better on conceptual questions than students in the IFN group, F(1, 302) = 12.21, p = .001. Both partial η² and Cohen’s d indicated a medium effect (η² = .033, d = .30).

We also conducted ANCOVA on the final course grades to determine if students in the IFN and IPN groups earned significantly different grades in the course. We designed the courses such that different types of classroom activities (in-class assignments, homework, etc.) received the same relative weight in the grading system to ensure that course grades across the two instructors were made up of the same types of assignments valued at the same worth. Results revealed that students assigned to IPN performed significantly better in the course, F(1, 304) = 6.21, p = .013, than students assigned to IFN. Both partial η² and Cohen’s d indicated a small effect (η² = .020, d = .19).

To determine the effect of type of notes on attendance, we conducted an independent samples t test on the random attendance probes taken throughout the semester. A research assistant unobtrusively collected attendance data for 20% of class periods on dates that we randomly selected from all possible class sessions. We then analyzed attendance data to determine differences between IFN and IPN on class attendance. Although students in the IPN group evidenced higher average attendance, this difference was not statistically significant, t(26) = 1.50, p = .14. On the exit survey, although the majority of students in both IFN and IPN groups indicated that the notes did not adversely affect their attendance in the class, more students in the IPN group endorsed the statement “The notes did not affect my attendance” (82.95%) than did students in the IFN condition (68.75%). This was a statistically significant difference, χ²(1, N = 297) = 11.56, p = .001, with φ = .20 indicating a small to medium effect. Although not directly related to our research questions, students also completed measures to determine the utility and social validity of the notes. In the IPN condition, 92.25% rated the notes as very or moderately useful, whereas 85.15% rated them similarly in the IPN condition. Students in both courses indicated that they regularly used the notes during lectures to help with note taking (78.75% in IFN; 70.45% in IPN). Neither of these differences were statistically significant, χ²(1, N = 297) = 2.66, ns; χ²(1, N = 297) = .57, ns. In addition, one question on the exit interview accessed possible contamination or “sharing” of the notes across sections to determine if there were threats to internal validity. The majority of students in both sections stated that they “never shared or borrowed notes from another section of PSY 101” (86.7% in IFN, 72.5% in IPN). This difference was statistically significant, χ²(1, N = 297) = 8.52, p = .004, with φ = .17 indicating a small to medium effect.

Discussion

The purpose of our study was to determine the differential effect of type of instructor-provided notes on
examination grades, final grades, and attendance. Contrary to Vandehey et al. (2005), our results provided some support for the hypothesis that partial notes led to better performance, particularly on cumulative conceptual questions and final grades. Although the effect was small, students assigned to the IPN condition performed significantly better on Exam 3 and on the cumulative final exam (Exam 4) than students in the IFN condition. No statistically significant differences emerged on the first two examinations. Students in the IPN group also earned significantly higher course grades than their counterparts in the IFN group. We did not find differential effects of note type on class attendance, although the data suggested that students in the IPN condition were more likely to self-report that having access to the notes affected their attendance.

The data on examination performance suggested that initially, type of notes did not lead to significant differences in exam scores. However, as the semester progressed, students receiving IPN outperformed students receiving IFN on examinations, including the final exam. There are several possible explanations for these findings. First, it is possible that as the semester progressed, because the students in the IFN condition had all the lecture material, they relied more on the notes and stopped engaging in other behaviors to facilitate learning (e.g., taking additional notes, attending to lecture, reading the textbook) whereas students in the IPN condition had to continue to engage in those behaviors to acquire the material to complete their notes. Perhaps at the beginning of the semester, all students engaged in adaptive study behaviors, but as the semester progressed, those behaviors declined for students who already had all the requisite material contained in their notes.

An additional explanation for these findings may be that students entering college for the first time may be ill-equipped in terms of study habits and, after performing poorly on the first few exams, may increase their study behaviors. Our data are consistent with this hypothesis; in all four courses, the exam scores were lowest for the first exam and gradually improved over the course of the semester. It may be that students in the IPN condition are better able to improve their study behaviors because of the encoding function of notes. For instance, when students finally do become a bit “panicked” by their low exam scores after the second test and begin studying, they perform better with the partial notes because they have encoded information better with these notes. Students assigned to the IPN condition also averaged higher final grades in the course, although these data were confounded by their better performance on the last two examinations in the course, relative to those in the IFN condition. In sum, these findings may suggest that instructors can facilitate active note taking by providing students with partial notes, because providing incomplete notes encourages students to elaborate on and encode material during the lecture. Consequently, student scores on examinations and, ultimately, the grade earned in the course may be affected.

An interesting finding in this study was the differential performance of students on conceptual questions on the final examination, although not on any of the previous examinations. We speculate that this differential performance on conceptual questions may relate to the active nature of the IPN condition relative to the amount of information presented in a cumulative final examination. The final examination required understanding of a large amount of material, and it may be that students with partial notes had been encoding information throughout the semester, rather than just memorizing notes and definitions. On a test that required knowledge of a large number of concepts, rote memorization was not feasible, so students who encoded the information by actively taking notes throughout the semester may have performed better because they had experienced better conceptual understanding.

Our study failed to find a significant difference in attendance between students assigned to IPN compared to IFN, although there was a trend toward significance. We assessed attendance via probes during 20% of the lectures; thus it is possible that the failure to find a difference was the result of low power to detect a difference between the two groups, rather than a lack of difference. A post hoc analysis of power at $\alpha = .05$ estimated power at .314 and suggested that at least 22 probes for each class would have been necessary to reach significance given the difference evidenced in our sample. In future studies, researchers may choose to assess attendance more regularly and rigorously to determine if, in fact, a difference exists in attendance across note conditions. Interestingly, students assigned to use IFN were significantly more likely to report that the notes adversely affected their attendance in class compared to those using IPN. Thus, student perception data indicated that attendance may have been negatively impacted by providing full Web notes. Although our empirical data failed to demonstrate this difference, our data do suggest that instructors should use caution in providing highly detailed notes to students if attendance in class is important.
In general, our study provided support for the encoding theory of notes, in that through the process of taking notes, students gain some mastery of the material. Additionally, the results suggested that instructors should be cautious in providing full notes to students, even when students often state that they prefer these notes. Providing notes to students may be desirable in terms of what they say they want, but appears to have an adverse effect, especially on conceptual or application types of outcome measures. If the intent of college teaching is to help students master the material and facilitate higher level learning, then providing full notes appears to be inhibiting these processes. Partial notes, therefore, may provide a nice balance in terms of providing students with some notes, which they report as helpful, and still requiring encoding and higher level processing of information, which will ultimately improve learning and performance.

Limitations and Recommendations

Future research may be able to address a variety of weaknesses in our study. Our sample was composed primarily of first-semester freshmen enrolled in a general education course, which, although providing a window into possible deficits in study-behavior repertoire, prevented analysis of how behaviors change as one progresses in college. Future researchers may consider examining upper level and major-specific courses to determine the effects of such notes on performance in different domains. Additionally, the sample was composed primarily of female students; although this is not atypical of psychology courses, it limits the generality of these findings.

This study was also limited by the cursory nature of some of the outcome data. As mentioned previously, future researchers should consider other means of assessing attendance to provide a more sensitive and more statistically powerful means to determine differences between groups. Additionally, this study used traditional outcome measures of tests and grades, with a small variation of examining performance on conceptual and definitional questions on those outcome measures. Future researchers may consider using other, perhaps more dynamic, measures of learning that recognize that differences may exist between types of outcome measures, including but not limited to conceptual, application, definitional, and analysis types of outcomes. A useful model for this type of analysis appeared in Neef, McCord, and Ferreri (2006).

Finally, data from the exit survey suggested that anywhere from 15% to 28% of students reported sharing notes with individuals from another section of introductory psychology. The nature of the exit survey prevented us from determining whether these students were sharing notes with individuals from the same instructor’s other section, thus contaminating the independent variable. Even though only a minority of students reported sharing notes, the students who did share notes may have inadvertently contaminated the integrity of the experimental assignment. Future researchers should recognize this potential problem with online access to instructional material. With advances in information technology, printing an extra copy of notes is much less arduous than other methods, such as making another photocopy of notes at the library. The ease of online access, with its availability 24 hr a day, may also contribute to other poor study behaviors, including cramming and last-minute completion of course material, which would be an additional fruitful area of study. It is our hope that this research serves as a jumping-off point for future research designed to both address limitations of this project and expand it to other areas of teaching in psychology.

References


**Notes**

1. This study was supported by a grant from the Faculty Teaching and Learning Center at Grand Valley State University.

2. Send correspondence to Tara L. Cornelius, Grand Valley State University Department of Psychology, Allendale, MI 49401; e-mail: cornelta@gvsu.edu.
We assessed the effects of using LearnStar™, an interactive, computer-based teaching tool, as an in-class exam review method. Students with higher LearnStar review scores had higher grades. Furthermore, students’ satisfaction ratings indicated that LearnStar reviews were more enjoyable and conducive to participation than traditional reviews. However, students who reviewed using LearnStar did not have significantly higher exam scores or course grades compared to students who had traditional reviews. Future research directions include measuring different aspects of students’ engagement in courses and examining the effects of using tools such as LearnStar for other in-class activities such as brief quizzes or polls.

Maintaining students’ interest as well as engaging them in the classroom are two challenges faced by instructors. For example, if students are not interested in the course material, they are less likely to come to class (Galichon & Friedman, 1985). Furthermore, activities that increase active participation have improved students’ performance (Narayan, Heward, Gardner, Courson, & Omness, 1990), and students report greater enjoyment of classes using active learning techniques (e.g., Middlecamp, 2003; Zaremba & Dunn, 2004; Zehr, 2004).

Instructors have used many techniques to increase students’ interest and participation in the classroom. For example, Boniecki and Moore (2003) increased students’ participation by using a token economy, whereas Butler, Phillmann, and Smart (2001) used short writing exercises followed by discussion to increase students’ active learning in a lecture setting. Recently, another area of interest has been the use of computer-based techniques to increase participation, interest, and learning outcomes. Technology-assisted instruction tends to be associated with increased student motivation, enjoyment, learning, and development (Forsyth & Archer, 1997); however, learning outcomes are not always superior in technologically assisted classes (DeBord, Aruguete, & Muhlig, 2004). Thus, although computer-assisted instruction appears to be associated with positive outcomes in general, the research in this area is inconclusive. Furthermore, the specific tools and techniques used vary widely, and the effect of any one tool cannot be generalized across studies.

Our study assessed the effects of LearnStar™, a computer-based, interactive trivia-style game, on students’ reactions and performance in two different undergraduate psychology courses (General Psychology and Abnormal Psychology). Our goal was to increase student achievement and satisfaction with these courses. We hypothesized that using LearnStar would achieve these goals by increasing students’ level of interest, participation, and class performance.

### Method

#### Participants

Three hundred seventy-seven undergraduate students from six introductory psychology and two abnormal psychology classes participated. The number of students in the introductory psychology sections ranged from 48 to 51 (149 total in LearnStar classes; 148 total in traditional classes), with the majority (70% to 80%) being freshmen. Both abnormal psychology sections had 40 students, the majority of whom were at or above junior level (75% in the LearnStar class; 87% in the traditional class). The majority of students in both courses were women (60.6% women and 39.4% men in introductory classes; 72.5% women and 27.5% men in abnormal classes). Scheduling constraints (e.g., moving the LearnStar keypads from one classroom to the next) determined which classes received LearnStar; thus, complete random selection was not possible.

#### Measures

Four exam scores and final semester grades were measures of class performance. In classes using LearnStar, students received a review score during each of the four review sessions based on the accuracy and speed of their responses to the questions. At the end of the semester, all students completed a five-item questionnaire assessing how beneficial they found the review sessions. The students responded to each statement on a 6-point Likert-type scale ranging from 1 (strongly disagree), 2 (disagree), 3 (somewhat disagree), 4 (somewhat agree), 5 (agree), to 6 (strongly agree). This forced-choice response format prevented students from choosing a neutral answer.
Procedure

Each class participated in a review session the class period before each exam. All sections had four review sessions. In half of the classes (three introductory psychology classes and one abnormal psychology class), the instructor led the class in a review session using LearnStar. At the beginning of these reviews, each student received a keypad and a code name to maintain confidentiality. The instructor displayed multiple-choice review questions on a large screen located in the front of the classroom, and students used their keypads to submit their answers. After a set amount of time, the correct answer appeared on the screen along with the number of people who had answered the question correctly. Students who answered correctly received points toward their LearnStar review score, with faster responses resulting in more points, whereas students who answered incorrectly lost points. In the traditional review classes, instructors used a standard review format covering the same topics as in the LearnStar classes but not necessarily asking the same questions or using a question-and-answer format.

Classes were standardized as much as possible across conditions. Each introductory psychology instructor taught one LearnStar and one traditional course, using the same lecture notes and textbook for both sections. Two instructors cotaught the abnormal psychology classes, with one instructor teaching both classes during the first half of the semester and the other instructor teaching both classes during the second half; the instructors used the same textbook and lecture notes for both classes. In addition, the class meeting times were counterbalanced as much as possible across conditions.

Results

Comparison of Exam Scores and Final Grades

We used a two-way ANOVA, with instructor and type of review as independent variables, to compare students’ final grades. The level of statistical significance was .05 for all analyses. LearnStar students’ final grades were not significantly different from regular-review students’ grades, $F(1, 361) = 0.99, p = .32$, nor was there a significant Review Type × Instructor interaction, $F(3, 361) = 0.37, p = .77$. In addition, the LearnStar students’ mean exam scores did not differ significantly from the traditional review students’ scores, $F(1, 369) = .69, p = .41$, nor did the Review Type × Instructor interaction have a significant effect on these scores, $F(3, 369) = 1.03, p = .38$.

Predicting Exam Scores and Final Grades

We tested LearnStar review scores for their ability to predict class performance. First, we predicted each exam score using a hierarchical regression model with instructor entered as the first predictor (to account for variance due to different grading standards) and the LearnStar review score entered as the second predictor. The overall model was significant for each exam (Exam 1: $R^2 = .52, p < .001$; Exam 2: $R^2 = .29, p < .001$; Exam 3: $R^2 = .29, p < .001$; Exam 4: $R^2 = .17, p < .001$). In all analyses except Exam 4, the LearnStar review scores predicted a significant proportion of the variance in their respective exam scores over and above the effects of instructor differences (Exam 1: $pr^2 = .37, p < .001$; Exam 2: $pr^2 = .32, p < .001$; Exam 3: $pr^2 = .25, p = .001$; Exam 4: $pr^2 = .15, ns$). Furthermore, we predicted final grades from LearnStar review scores. The overall model was significant, indicating that the optimal linear combination of all four LearnStar review scores accounted for a significant portion of the variance in final grades, $R^2 = .18, p = .014$.

Satisfaction Analysis

See Table 1 for means, standard deviations, significance levels, and effect sizes for all analyses discussed in this section. The mean satisfaction score, created by averaging ratings on the five statements assessing how beneficial students found the reviews, was significantly higher for the students participating in the LearnStar reviews compared to those participating in the traditional reviews. However, the difference in these mean scores was small (less than half of a point), and the partial eta-squared statistic indicated that this effect was negligible (ex-

### Table 1. Mean Ratings on Satisfaction Statements

<table>
<thead>
<tr>
<th>Statement</th>
<th>LearnStar™ M</th>
<th>LearnStar™ SD</th>
<th>Traditional M</th>
<th>Traditional SD</th>
<th>$F(1, 254)$</th>
<th>$\eta^2_{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Compared to reviews in other classes, I participated more during these reviews.</td>
<td>5.03</td>
<td>1.17</td>
<td>4.03</td>
<td>1.38</td>
<td>39.25**</td>
<td>.13</td>
</tr>
<tr>
<td>2. Compared to reviews in other classes, LearnStar [or the instructor] made reviews more interesting.</td>
<td>5.15</td>
<td>1.23</td>
<td>4.48</td>
<td>1.17</td>
<td>19.76**</td>
<td>.07</td>
</tr>
<tr>
<td>3. Overall, I enjoyed LearnStar [or the instructor’s reviews].</td>
<td>5.04</td>
<td>1.23</td>
<td>4.60</td>
<td>1.12</td>
<td>9.02*</td>
<td>.03</td>
</tr>
<tr>
<td>4. I found LearnStar [or the instructor’s feedback] beneficial for reviews.</td>
<td>4.80</td>
<td>1.24</td>
<td>4.80</td>
<td>1.02</td>
<td>&lt; 0.01</td>
<td>—</td>
</tr>
<tr>
<td>5. The review helped me prepare for exams.</td>
<td>4.53</td>
<td>1.35</td>
<td>4.45</td>
<td>1.31</td>
<td>0.27</td>
<td>—</td>
</tr>
<tr>
<td>Mean satisfaction score</td>
<td>4.91</td>
<td>1.13</td>
<td>4.47</td>
<td>1.02</td>
<td>10.69*</td>
<td>.04</td>
</tr>
</tbody>
</table>

*Note. Possible range of scores was from 1 (strongly disagree) to 6 (strongly agree). *p < .005. **p < .001.
Discussion

Instructors are often faced with the challenge of inducing student interest in course material. Furthermore, instructors must find ways to maintain this interest. Our study examined the use of LearnStar, an interactive computer technology tool, to increase interest and achievement in the classroom. Our results suggest that students in the LearnStar classes did not achieve significantly different final grades or exam scores compared to students in the traditional-review classes. Thus, using this type of tool for exam review may not be an effective way to improve student performance.

Regression analyses indicated that students who performed better during LearnStar reviews also performed better on exams and in the class as a whole. Thus, performance on LearnStar reviews predicted student performance. However, we were unable to compare this predictive ability to the predictive ability of traditional reviews, as the traditional reviews did not yield scores assessing knowledge of the topics covered. Furthermore, the amount of variance in final grades accounted for by LearnStar review scores, although significant, was small (18%), indicating that knowledge of review topics was a relatively minor predictor of final grades. Nonetheless, the review scores' predictive ability, especially their ability to predict exam scores, could give instructors valuable information, such as allowing them to gauge student knowledge of topic areas and helping them determine which areas need more discussion.

Although the LearnStar reviews did not appear to improve student performance, student evaluations showed that students preferred this type of interactive tool when reviewing for exams. Specifically, students in the LearnStar reviews reported finding the reviews to be more beneficial than those students participating in the traditional reviews. Compared to previous classes students had taken at the university, students in the LearnStar review classes reported that they found the reviews more interesting, participated more in class during the reviews, and enjoyed the reviews more overall. However, the low effect sizes for these results indicate that the type of review was not the most important contributor to student satisfaction.

These results may have been influenced by several uncontrolled variables. One such variable may have been class attendance. Specifically, some students attended class only on exam days and did not participate in the reviews. Thus, they did not benefit from either type of review. Future studies might attempt to control for attendance to assess better the effects of a tool such as LearnStar. Another area for research is the potential use of such technology for in-class activities other than reviews. As yet, few researchers have examined possibilities in this area. Instructors or researchers could use LearnStar or a similar tool to increase engagement during lectures, such as by using it to give brief quizzes on the information being covered or to conduct opinion polls on controversial topics.

In summary, LearnStar scores predicted class performance, and its use was associated with small increases in student satisfaction with and participation in reviews. However, using LearnStar did not improve students’ grades. These results suggest that although computer-based learning aids may increase student participation, they may not be an effective way to improve academic performance.

References


Notes

1. This research was supported by Grant NSF 0311617.
2. Portions of this article were presented at the 12th Annual Southwest Teachers of Psychology Conference, Seguin, TX, in November 2004.
3. We thank Robert Blake and the following teaching assistants for their assistance with this research project: Nidal Karim, Jason Frizzell, and Elizabeth Garza. We also thank Angela Lee and Jacqueline McNeil for their assistance in collecting the data.
4. Information on LearnStar™ can be obtained at www.LearnStar.com or via e-mail at sales@LearnStar.com.
5. Send correspondence to Joy R. Pemberton, Department of Psychology, Texas Tech University, Lubbock, TX 79409; e-mail: joy.reeves@ttu.edu.
We developed a computerized game designed around principles shown to improve classroom performance and experience: multimedia, practice testing, vivid instructional techniques, and in-class participation. Students compete as 2 teams, and the game plays like Jeopardy meets Hollywood Squares with learning outcomes. Our evaluation indicated that the game was effective in small to larger classes (28 to more than 50 students). We present data supporting the game’s effectiveness, and we discuss students’ reactions and data concerning self-reported effectiveness and perceived value.

Students who become active participants in learning tend to outperform students who do not in terms of exam scores, classroom participation, and overall course grade (Ainley, 1993; Finn & Rock, 1997; McManus, Dunn, & Denig, 2003; O’Sullivan & Copper, 2003; Skinner, Wellborn, & Connell, 1990). Researchers have proposed a variety of approaches and techniques for engaging students. For example, techniques range from simply integrating student-reported topics of interest into lectures (Buskist & Wylie, 1998) to active role-playing events (DeNeve & Heppner, 1997) and in-class simulations and collaborative teaching opportunities (Bernstein, Scheerhorn, & Ritter, 2002). Generally, there appears to be a positive correlation between involvement and achievement (Ainley, 1993; Finn & Rock, 1997; Skinner et al., 1990).

We attempted to develop an effective classroom tool that both engages students and improves their motivation to learn the material. Specifically, we designed an interactive review game called PsychOUT! that takes advantage of classroom technologies. The pedagogy we emphasized in the design of this game was the review exam. Because questions and answers are easy to adapt from resources that often accompany the instructor’s text or other course content, the potential use of the game is not limited to specific course topics.

Our goals were (a) to assess students’ perceived effectiveness of the game in affecting their exam performance and study habits and (b) to determine whether playing the game actually affected students’ exam performance. Using both open- and closed-ended questions, we asked students to rate aspects of the game, lectures, and the class. In addition, we surveyed students for input as to the potential problems and benefits of using the game in a classroom setting. We also compared exam scores across two sections of a general psychology course for students who did and who did not participate in the game.

Method

Participants

Our evaluation of the game took place over the course of two semesters in four classrooms: Three were undergraduate general psychology courses (ns = 115, 74, 32), and the fourth was an upper level social psychology course (n = 28). We obtained only survey data from the social psychology course and one of the general psychology courses (n = 32). We used the remaining general psychology courses to examine test scores (total n = 189). Of these, we were able to survey only the smaller class (n = 74). We did not require any students to participate in the PsychOUT! review sessions, although we did offer extra credit to students who opted to come to class and play the game (123 students opted to play). Students who did not attend the class when the game was played made up the “nonplayers” comparison group (n = 66). Also, we added the extra-credit points to final course averages rather than to the exam scores we planned to analyze.

Materials and Apparatus

PsychOUT! derives its name from one of the play options. Students may occasionally opt to try and answer a question with a bluff to “psych out” the opposing team and win extra points. Standard play, however, is simply a matter of each
team taking a turn at choosing a question topic and difficulty level and then answering it correctly for points.

The game requires access to a computer (PC running Windows 95 or later) connected to a color projector and speakers (although sounds are not essential, they do add to the experience). A $5 \times 5$ array of boxes (buttons) makes up the basic gameboard display. This arrangement allows for 25 question–answer opportunities per round (up to three rounds can be played without restarting the program). Above each column of five buttons is a question-category description. Inside each of the five buttons is a value that represents the number of points available for the playing team to attempt should they select that box. The master of ceremonies (MC) clicks the mouse on whichever button the playing team selects, and a question–response opportunity occurs. That is, a question appears (e.g., “Condition in which stimulation in one sensory modality arouses imagery in a different modality.”) above a color animation that depicts the amount of time remaining for an answer (beginning with 30 sec). Once the team provides a response, the MC stops the timer and the correct answer appears (e.g., “Synesthesia”). It is up to the MC to assess whether the team answer matches the answer displayed by the game. A “partial points” option is available if the MC believes the answering team deserves some (rather than none) of the points. Play alternates between opposing teams until the round is over. We found that 25 questions take about 25 min to complete, so a two-round game (50 questions) is best for a 1-hr class, whereas the three-round game (75 questions) would likely work well for a 90-min class.

Design

We used a $2 \times 2 \times 4$ mixed design for the analysis of exam performance. Consequently, we treated game presentation (before Exam 1 and before Exam 2) and participation (played and did not play) as between-subject variables, and exam (1, 2, 3, and 4) as a within-subjects variable. The dependent measure was students’ scores for each of the four exams.

Procedure

Students played the game during the class meeting before an exam (i.e., 2 full days before the exam). We analyzed exam performance across two semesters of general psychology. Students in the spring semester played the game before the first exam, and fall-semester students played the game before the second exam. In addition, because not all of the students chose to play the game, we were able to compare students who did and who did not participate in the game.

For the evaluation of PsychOUT!, we surveyed students at the end of the semester (two general psychology and one social psychology) along with the course evaluations. The PsychOUT! survey contained nine 7-point Likert scale questions (where 1 indicated a negative evaluation for that question and 7 indicated a positive evaluation) and six open-ended questions. The Likert questions assessed the degree to which students believed the game was an effective or valuable learning tool. The open-ended questions allowed for a greater range of expression from students (whether they played games in other classes, suggestions for improvements, effects of playing on study strategy, etc.).

Results

We distributed the PsychOUT! surveys on the last day of classes, and not all students were present. Therefore, of the 134 students who could have completed the survey, 90 did so (67%). Students’ mean self-reports were high for all questions (i.e., all means were greater than 5.4 on the 7-point Likert scale). To assess participants’ ratings of the game, we compared students’ mean responses to the neutral Likert response point (i.e., 4). In every case, response means were significantly greater (i.e., all $t > 9.60, p < .01$) than the neutral point (see Table 1). We did not perform any analyses on the responses to the open-ended questions because our goal for those questions was simply to collect insights and responses about the game rather than to test any specific hypotheses.

For the analysis of class performance, we performed a 2 (game presentation) $\times$ 2 (participation) $\times$ 4 (exam) mixed factor ANOVA on students’ exam scores. We found a significant main effect for participation, $F(1, 185) = 16.93, p < .01$ ($\eta^2 = .092$), in which students who participated in PsychOUT! averaged higher exam scores ($M = 72.7, SD = 15.9$) than students who did not participate ($M = 61.9, SD = 14.5$).

We found a significant interaction Participation $\times$ Exam interaction, $F(3, 555) = 3.84, p < .05$ ($\eta^2 = .021$). Because the critical comparison involved when students first played the game (i.e., before Exam 1 vs. before Exam 2), we reanalyzed the data excluding students’ scores from the last two exams. This analysis yielded a significant three-factor interaction, $F(1, 185) = 5.74, p < .02$ ($\eta^2 = .031$). That is, all PsychOUT!-played conditions resulted in higher scores than nonplayed conditions. When students played PsychOUT! before their first exam, their average exam score was better than that of the students who did not play. However, in the class that did not play the game before the first exam, scores were relatively poor. Once both classes played the game (i.e., by Exam 2) students who played PsychOUT! averaged higher exam scores than students who did not play. We present the outcome data from this analysis, along with the data for Exams 3 and 4, in Table 2.

Discussion

In our experience, and based on responses to the open-ended survey questions, some students arrive to play
PsychOUT! with greater confidence in their knowledge of the course content than when they finish playing the game. Numerous responses to the surveys strongly indicated that it was the game experience that revealed to students the inadequacy of their study habits or the incompleteness of their content knowledge. As a result, students claimed that they reevaluated their study methods to better prepare themselves for exams. Indeed, students who played PsychOUT! generally achieved higher exam scores than those who did not.

Ultimately, our view is that using review tools such as PsychOUT! in the classroom helps to emphasize the metacognitive aspects of learning. Specifically, we hypothesize that it is through the dynamic conditions of “game play” that metacognitive insights occur compared with typical classroom review conditions. The typical classroom review is likely to be less engaging and, based on our anecdotal experiences, sometimes supports post hoc rationalizations that protect students’ sense that recognition familiarity is sufficient for exam preparedness. Also, by writing questions that emphasize recall of information over recognition, many students will likely experience the cost of having studied only for familiarity in preparing for game play.

Despite the success of PsychOUT! in our research, we are nonetheless hesitant to make strong claims about the generalizability of the results. Our concern has to do with the review content. We do not believe that our results could support the broad claim that use of this or any other game guarantees improvements among students who play. It is important to remember that there are two critical components to any review tool: the tool (or delivery method) and the review material presented. In our study, we held review material relatively constant. Therefore we do not feel comfortable generalizing the findings to conditions in which review content is free to vary. Because the effectiveness of any review tool must depend a great deal on the review material used, we can easily envision a situation where game play might actually result in worse exam performance for students than not having played at all. For example, consider the implications should an instructor select questions for the game that are much easier than those to appear in the upcoming exam. Students might erroneously conclude that because they were successful playing the game they will be equally successful on the coming exam. They may, therefore, be less likely to study for the exam than had they not observed the game at all.

Table 1. Means, Standard Deviations, and Effect Sizes (d) for Ratings of the 9 Likert Scale Questions in Comparing Each Mean Response to Neutral Point

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Do you believe that playing PsychOUT! helped you to learn material for class?</td>
<td>5.54</td>
<td>1.27</td>
<td>1.21</td>
</tr>
<tr>
<td>2 To what extent do you believe that playing PsychOUT! is a valuable use of class time?</td>
<td>5.68</td>
<td>1.26</td>
<td>1.33</td>
</tr>
<tr>
<td>3 Would you say that in-class games like PsychOUT! are effective methods for teachers to use in order to get students to learn course-related information?</td>
<td>5.89</td>
<td>1.28</td>
<td>1.48</td>
</tr>
<tr>
<td>4 In your opinion, is the use of multimedia technology in the classroom (e.g., PowerPoint, video clips, games like PsychOUT!) something that actually improves most, if not all, students’ performance in the course?</td>
<td>5.98</td>
<td>0.91</td>
<td>2.18</td>
</tr>
<tr>
<td>5 In your opinion, is the classroom experience made better by a teacher’s use of multimedia technology during lectures (e.g., PowerPoint, video clips)?</td>
<td>5.99</td>
<td>0.99</td>
<td>2.01</td>
</tr>
<tr>
<td>6 Do you feel that playing PsychOUT! serves as a good review before an exam?</td>
<td>5.97</td>
<td>1.33</td>
<td>1.48</td>
</tr>
<tr>
<td>7 How vivid (freshness of experience, memorable, etc.) was it to play the PsychOUT! game in class?</td>
<td>5.46</td>
<td>1.35</td>
<td>1.08</td>
</tr>
<tr>
<td>8 The idea of playing PsychOUT! is based on findings that practice testing improves student performance on exams. To what extent do you believe that this game was effective in improving your performance on exams?</td>
<td>5.50</td>
<td>1.16</td>
<td>1.29</td>
</tr>
<tr>
<td>9 The way in which PsychOUT! is played is based on findings that class participation improves student performance. Given that playing this game requires everyone to participate, to what extent do you believe that playing PsychOUT! made you feel that you were participating in class?</td>
<td>5.41</td>
<td>1.39</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Note. All ps < .01.

Table 2. Descriptive Statistics for Conditions

<table>
<thead>
<tr>
<th>Game Presentation</th>
<th>Exam</th>
<th>Played PsychOUT!</th>
<th>Did Not Play PsychOUT!</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Before Exam 1</td>
<td>1 76.2</td>
<td>14.1</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>2 71.7</td>
<td>16.5</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>3 77.8</td>
<td>15.2</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>4 73.2</td>
<td>14.2</td>
<td>63</td>
</tr>
<tr>
<td>Before Exam 2</td>
<td>1 65.5</td>
<td>16.0</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>2 71.0</td>
<td>18.4</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>3 75.2</td>
<td>16.7</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>4 70.9</td>
<td>16.2</td>
<td>60</td>
</tr>
</tbody>
</table>
As a final note, we would like to share that some of the less satisfied students believed that they would have been better served playing the game in paired groups (or triads, etc.), rather than being put “on the spot” during game play. The number of students contributing a response during game play is obviously going to be a function of the class size. Larger classes (e.g., 40 or more students) might be ideally suited for this variation of game play. However, the risk to larger answer groups is an increased likelihood of social loafing because students would have proportionally less opportunity to actively participate.

References


Notes

1. Portions of this research were presented at the 26th Annual National Institute on the Teaching of Psychology, St. Petersburg Beach, FL, in January 2004.
2. Appreciation and thanks are extended to Allen Liss for his assistance. The game engine is currently distributed not-for-profit by drspeg Software, Inc. Address inquiries pertaining to cost and availability to drspeg@hotmail.com. Please put the word “PSYCHOUT” in the subject line.
3. Send correspondence to Stephen T. Paul, Department of Social Sciences, Robert Morris University, 6001 University Boulevard, Moon Township, PA 15108; e-mail: paul@rmu.edu.

Call for Applications for E-Publications Editor-in-Chief for the Society for the Teaching of Psychology

The Society for the Teaching of Psychology (STP) is searching for an E-Publications Editor-in-Chief to serve a 2-year term beginning officially on January 1, 2008. This is a new position for STP and the initial short-term length is to evaluate the work load and the demand for future e-books. After this time period, it is expected that the term length will be increased to 6 years. The editor-in-chief will be appointed by August 2007 and begin working with the STP Publications Committee, Internet editor, and director of the Office of Teaching Resources in Psychology (OTRP) to develop and implement plans for STP's e-publishing efforts by early 2008. The STP will provide $3,000 per year for a one-course reduction in the successful applicant’s teaching load.

Since 2001, STP has published three e-books plus a four-volume e-series of books from PsychTeacher's E-xcellence in Teaching essays. Plans are currently underway for two more e-books and a fifth volume of E-xcellence in Teaching essays. All of STP's e-books are published on its home page (http://teachpsych.lemoyne.edu/teachpsych/div/divindex.html) and are downloadable chapter-by-chapter or as entire books in rtf, doc, and html formats.

The editor's responsibilities include the following:

- reviewing, in consultation with the chair of the publications committee, the director of OTRP, the Internet editor, and members of the consulting editorial board, all e-proposals for their suitability for publication;
- reviewing, editing, and formatting all e-publication materials;
- making an annual report to the STP Publications Committee, and
- maintaining communication with the executive committee of STP as appropriate.

Applicants must be members of the Society, be familiar with the editorial and publication process in general, be familiar with e-publication processes and formatting documents for Internet use or demonstrate a willingness to learn formatting skills, have excellent organizational skills, have excellent communication skills, and be able to commit sufficient time to the position.

Applicants should send a cover letter detailing their relevant experience and qualifications for the position. They should also include a current version of their vita and the names and contact information of three people who can speak to their qualifications for the position.

All materials must be received by March 15, 2007. Please send all application materials to Stephen F. Davis, Chair, STP Publications Committee, at davis122@cox.net using attachments in either Word or WordPerfect.
Promoting Active Learning Using Individual Response Technology in Large Introductory Psychology Classes

Christopher R. Poirier
Stonehill College

Robert S. Feldman
University of Massachusetts, Amherst

Individual response technology (IRT), in which students use wireless handsets to communicate real-time responses, permits the recording and display of aggregated student responses during class. In comparison to a traditional class that did not employ IRT, students using IRT performed better on exams and held positive attitudes toward the technology. IRT appears to be a promising technology for increasing active learning in the classroom and enhancing students’ mastery of course content.

Research suggests that active learning exercises can be powerful teaching tools to facilitate learning (Butler, Phillmann, & Smart, 2001; Yoder & Hochevar, 2005). However, enhancing lectures with active learning exercises can be challenging for instructors teaching at large institutions because instructors often teach in lecture halls with hundreds of students. Although the large lecture hall format makes efficient use of space and faculty, students’ mastery of course content may be compromised.

Individual response technologies (IRTs) allow instructors of large classes to implement student-centered activities, which offer students an opportunity to be active learners. Using IRT, each student uses a wireless personal handset, similar to a television remote control, to transmit answers to multiple-choice questions. When a student presses a button, the handset transmits a signal to a receiver and then to a personal computer, which records and aggregates the input and displays the information in real time, allowing students and instructors to discuss the data.

There are several potential advantages of this technology (Boyle & Nichol, 2003; d’Inverno, Davis, & White, 2003; Draper, Cargill, & Cutts, 2002; Wit, 2003). First, IRT allows all students to participate in demonstrations and activities and share their beliefs or knowledge, regardless of class size. In addition, because students’ responses are anonymous to other students, they are able to contribute without fear of public speaking or being ridiculed for making mistakes. Third, instructors can use IRT to assess students’ knowledge and receive immediate feedback on the lecture, allowing instructors to adjust their lecture to the audience. Finally, increasing student engagement and involvement in the class has the potential to lead students to expend more effort and therefore to achieve at higher levels.

There also are potential drawbacks to IRT use. For one, the technology may be an additional expense for students (about $15 per student) or institutions (some departments purchase a set of handsets for class use). Also, recording responses using the lower priced infrared system may take class time (about 5 min per session with 500 students). However, the newer radio frequency systems allow faster recording of responses, greatly reducing the amount of time required for an activity. Finally, similar to the use of other technologies, instructors may encounter occasional technical difficulties, disrupting demonstrations and activities or the flow of a lecture.

To determine the general efficacy of IRT, we examined exam grades and course evaluations from two
Table 1. Student Ratings of Individual Response Technology

<table>
<thead>
<tr>
<th>Item</th>
<th>N</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I was better prepared for class because I would need to answer IRT questions.</td>
<td>379</td>
<td>10.6%</td>
<td>34.8%</td>
<td>35.6%</td>
<td>19.0%</td>
</tr>
<tr>
<td>2. I learned more in class because IRT questions made me apply what the professor taught during the class.</td>
<td>383</td>
<td>15.1%</td>
<td>50.9%</td>
<td>23.2%</td>
<td>10.7%</td>
</tr>
<tr>
<td>3. Using IRT was fun and made the class more enjoyable.</td>
<td>384</td>
<td>28.1%</td>
<td>42.5%</td>
<td>18.5%</td>
<td>10.9%</td>
</tr>
<tr>
<td>4. Answering IRT questions encouraged me to talk to other students.</td>
<td>383</td>
<td>18.3%</td>
<td>49.1%</td>
<td>25.1%</td>
<td>7.6%</td>
</tr>
<tr>
<td>5. IRT did not significantly improve my learning in this class.</td>
<td>382</td>
<td>16.0%</td>
<td>38.7%</td>
<td>35.9%</td>
<td>9.4%</td>
</tr>
<tr>
<td>6. Using IRT was a waste of time.</td>
<td>387</td>
<td>14.5%</td>
<td>18.1%</td>
<td>38.0%</td>
<td>27.7%</td>
</tr>
</tbody>
</table>

large comparable introductory psychology courses, one that used IRT technology and one that did not. We hypothesized that the opportunity to participate in the IRT demonstrations would be enjoyable and promote more positive attitudes toward the course and higher exam performance.

Method

Participants and Procedure

A total of 865 undergraduate students, enrolled at a large state university, participated. Students enrolled in a traditional course (n = 418) or a course that incorporated IRT into lectures (n = 447). The courses met back to back, and students had no knowledge of which course used IRT until they actually enrolled. Both courses met for 75 min twice per week with the same instructor.

The introductory courses consisted of a combination of lecture, video, discussion, in-class writing, and exams. In addition, students in the IRT section participated in one IRT activity per week that included a combination of individual and group activities and demonstrations. Each IRT activity concluded with one and three multiple-choice questions to measure students’ understanding of content or attitudes about a topic. For instance, after discussing the concept of classical conditioning, the instructor worked through an example by asking several questions (e.g., What is the unconditioned stimulus?). In other cases, the instructor measured students’ attitudes and opinions before or after presenting content. For example, after presentation of material on the ethics of the use of animals in research, students answered a question about when they would find such research acceptable. For all sessions, the IRT software aggregated and displayed the results, allowing students to receive immediate feedback and compare themselves to others in the class. In turn, this information provided an opportunity for additional class discussion. In all cases of IRT use, the instructor presented material that he also presented in the non-IRT section.

Measures

To determine that the students were equivalent in the two courses, we compared high school grade-point average (GPA) and SAT scores. In addition, to see if IRT use was related to overall course performance and evaluations, we analyzed students’ exam grades and their ratings of the course and the use of IRT. The course evaluation (a standard measure obtained in all university courses) consisted of eight items, most of which students rated on a scale from 1 (almost always) to 5 (never). The IRT evaluation consisted of six items, which students rated on a scale from 1 (strongly agree) to 4 (strongly disagree; see Table 1).

Results

High School GPA and SAT Scores

We analyzed high school GPAs and SAT scores to determine that the two samples of students were equivalent prior to the experiment, and we found no statistically significant differences (p > .05).

Exam Grades

To test whether IRT use was related to students’ overall exam performance, we averaged students’ four
exam grades and performed an independent-groups t test comparing the grades. The test revealed that students in the IRT course performed significantly better ($M = 84.03\%$, $SD = 7.54\%$) than students in the traditional course ($M = 82.72\%$, $SD = 7.64\%$), $t(863) = 2.54$, $p = .01$, Cohen's $d = 0.17$.

Course Evaluations

There were no statistically significant differences in students’ course evaluation ratings ($p > .05$); students in both courses gave generally positive feedback. Furthermore, students reported generally positive attitudes toward IRT use (see Table 1). Over 66% of students believed that they learned more during class because IRT allowed them to apply what the professor taught. In addition, 67% of students agreed that IRT encouraged them to talk to other students, and over 70% of students thought that IRT was fun and made the class more enjoyable. However, the feedback regarding the use of IRT was not consistently positive. Although students reported learning more during class because of IRT, nearly half of the students reported that IRT did not significantly improve their classroom learning.

Discussion

This study examined the efficacy of using IRT in a large introductory psychology course. The results suggest that IRT can be a useful active learning tool in a large course: Most students gave positive feedback regarding IRT use. In addition, although IRT sessions made up less than 10% of class time, students earned higher exam scores in the IRT section. Perhaps the use of IRT disrupted the monotony of lecturing and allowed students to pay attention longer, causing them to learn more key concepts.

Attitudes toward IRT use were not uniformly positive. Anecdotally, some students noted to the instructor that the use of IRT replaced what they saw as more valuable lecture time. In addition, although students reported learning more in class because IRT made them apply what the instructor taught, approximately half of the students believed that IRT use did not significantly increase the amount that they learned in class, suggesting that some students believed that IRT had a minimal effect on learning, which is consistent with the small effect size.

Apparently, reactions to IRT varied substantially from one student to another, suggesting the importance of future research that examines individual differences in student acceptance of technology and its impact on performance. Future researchers should identify the factors that help students willingly embrace the use of IRT technology and investigate how to make use of IRT most effectively from a pedagogical point of view. Still, the results suggest that IRT provides significant benefits. IRT appears to be a promising technology for increasing student involvement and has the potential to enhance learning.

References


Notes

1. Portions of this article were presented at the annual meeting of the American Psychological Society in Los Angeles, May 2005.
2. Send correspondence to Robert S. Feldman, Department of Psychology, University of Massachusetts, Amherst, MA 01003; e-mail: feldman@psych.umass.edu.
Benefits of Electronic Audience Response Systems on Student Participation, Learning, and Emotion

Jeffrey R. Stowell and Jason M. Nelson
Eastern Illinois University

We compared an electronic audience response system (clickers) to standard lecture, hand-raising, and response card methods of student feedback in simulated introductory psychology classes. After hearing the same 30-min psychology lecture, participants in the clicker group had the highest classroom participation, followed by the response card group, both of which were significantly higher than the hand-raising group. Participants in the clicker group also reported greater positive emotion during the lecture and were more likely to respond honestly to in-class review questions.

Increasing student participation is one of many strategies that might lead to improved student learning. To increase student participation, instructors can use “active student responding” methods (Heward, 1994). One method is the use of paper response cards, with possible answer choices such as True/False or A/B/C/D that students can hold up when the instructor poses a question. Response card usage has resulted in increased student participation and improved academic performance (Gardner, Heward, & Grossi, 1994).

More recent active student responding methods incorporate technology that allows students to send their responses electronically from hand-held keypads (clickers) to a receiver attached to a computer. The computer instantly tallies and graphically displays student responses on the computer screen.

Our study examined the impact of clickers on student participation and academic performance. Additionally, we investigated the effect of clickers on academic emotions “directly linked to academic learning, classroom instruction, and achievement” (Pekrun, Goetz, Titz, & Perry, 2002, p. 92). Academic emotions are related to important processes associated with academic performance including metacognition, strategy usage, and working memory functioning (Pekrun, Elliot, & Maier, 2006). Therefore, it is important to know not only how instructional techniques impact knowledge acquisition but also academic emotion (Nelson & Manset-Williamson, 2006).

As noted by others, one of the main differences between clickers and other student responding methods is the former’s allowance of anonymous responding. Students privately press a button on their clicker, whereas other forms of active student responding (e.g., hand-raising, response cards) require conspicuous behaviors. Although a direct relation between such anonymity and academic performance is unknown, its potential impact on student participation and academic emotions is more apparent. For example, students who have tendencies toward introversion might be more willing to participate and might experience less negative academic emotions (e.g., anxiety, shame) when using an anonymous responding method.

Other major differences between electronic keypads and other student responding methods include more immediate feedback to instructors and the graphic display of polling results. Whether these differences result in improved academic performance and more adaptive academic emotions is unknown. The purpose of our study was to examine whether the use of clickers in an undergraduate setting would result in greater learning, participation, honesty of student feedback, and more positive academic emotions than other methods of student responding.

Method

Participants

One hundred forty undergraduate students enrolled in introductory psychology classes at a public regional institution in the Midwest participated. Of those, 70% were women and 77% were in their first year of college.

We recruited students from an introductory psychology research pool and awarded 1 hr of research participation credit toward their 4-hr requirement. As motivation to put forth effort on the learning assessment measures, we awarded entries into a drawing for a $25 gift certificate to a local store, based on their postlecture quiz performance.
Materials

Surveys. The Academic Emotions Questionnaire (AEQ; Pekrun et al., 2002) measures emotions over time in academic settings such as a lecture or examination. Twenty-three items measure enjoyment, hope, anger, anxiety, and hopelessness experienced before the event (AEQ-Before). Forty-two items measure enjoyment, hope, pride, anger, anxiety, shame, hopelessness, and boredom experienced during the event (AEQ-During), and 15 items assess enjoyment, pride, anger, shame, and hopelessness experienced after the event (AEQ-After). Participants rated individual items on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). We asked participants specifically about their emotions concerning the classroom lecture.

Response cards. We made response cards by laminating sheets of white paper, with the choices of 1 and 2 printed in 80-point Arial font at the top and bottom, respectively. A second response card had the choices 3 and 4 printed on it. After the instructor posed a multiple-choice question and said, “Answers please,” students simultaneously raised their selected response card to show the instructor their answers.

Clickers. The clickers were approximately the size of a credit card and sent an infrared signal to a receiver attached to a computer. The polling software (TurningPoint, 2006) was seamlessly integrated into Microsoft PowerPoint and graphically displayed the percentage of individuals responding to each answer choice.

Depending on the time participants elected to participate (3:30 p.m. on a Tuesday or Thursday of 2 consecutive weeks), we assigned them to one of the following groups:

1. Standard lecture (n = 34). Throughout the lecture, the instructor spontaneously posed open-ended informal questions to students about the lecture material, calling on students who raised their hands.
2. Review questions: Handraising (n = 35). In addition to the informal questions, participants in this group received the closed-ended formal review questions (posed in a multiple-choice format) during the lecture. Students gave their answers to the review questions by raising their hands when asked how many thought each option was correct.
3. Review questions: Response cards (n = 36). Students in this group used response cards to indicate their answer to the formal review questions.
4. Review questions: Clickers (n = 35). Students used clickers to indicate their answers to the formal review questions.

Procedure

After providing informed consent, participants completed the AEQ-Before survey. Once they returned the survey, we told them they would be listening to a lecture on a topic in psychology and asked them to act as if they were in their regular psychology class by taking notes, asking questions, and making comments. The first author gave a 30-min introductory-level lecture on the organization of the nervous system, parts of the neuron, and how neurons work.

On different days, each group viewed the PowerPoint lecture in the same 50-seat classroom equipped with a computer and presentation software. An LCD projector displayed the computer monitor image on a white screen at the front of the classroom. From a front corner of the room, a camera videotaped the class.

Except for the standard lecture condition, we formally presented seven multiple-choice review questions, distributed throughout the lecture. After showing a formal review question on the screen, the instructor read aloud the question and four possible answers. After giving students a chance to indicate their answers in one of the ways described previously, the instructor stated the correct answer and continued with the lecture. The first author and an independent rater coded student participation and correctness of responses to the formal review questions from the videotapes. Tallies of participation and correctness by each rater were in high agreement, as indicated by an intraclass correlation coefficient of agreement of .98. In cases where there was a discrepancy, we used the mean values of the two raters for data analyses.

We counted the number of students who visibly provided an answer, the number of those who gave correct answers, and the number of visible nonresponders. We calculated formal participation rates as the number of responders divided by the total number of students.

We measured informal participation by counting the number of open-ended questions posed spontaneously by the instructor such as, “What branch of the nervous system is responsible for the fight or flight response?” We counted the total number of responses by the students to these questions and divided this by the number of informal questions asked to estimate an informal rate of participation.

After each lecture condition, participants completed the AEQ-During and the AEQ-After
Table 1. Participation Rates and Learning Performance by Assessment Technique

<table>
<thead>
<tr>
<th>Group</th>
<th>Informal Participation Ratio*</th>
<th>% Formal Participation</th>
<th>% Correct, Formal Review Questions</th>
<th>% Correct, Postlecture Quiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard lecture</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>57&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Hand-raising</td>
<td>1.20</td>
<td>76&lt;sub&gt;a&lt;/sub&gt;</td>
<td>98&lt;sub&gt;a&lt;/sub&gt;</td>
<td>60&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Response cards</td>
<td>1.21</td>
<td>97&lt;sub&gt;b&lt;/sub&gt;</td>
<td>92&lt;sub&gt;b&lt;/sub&gt;</td>
<td>52&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Clickers</td>
<td>1.11</td>
<td>100&lt;sub&gt;b&lt;/sub&gt;</td>
<td>82&lt;sub&gt;d&lt;/sub&gt;</td>
<td>60&lt;sub&gt;d&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Note. Means with different subscripts are significantly different from each other at p < .05 in a Tukey post-hoc comparison.

*Ratio of number of responses received to the number of informal questions asked.

surveys. Next, they completed a 10-item quiz based on the lecture. Finally, participants completed a survey about demographic information, preexisting knowledge of the lecture topic, and a five-item evaluation of the classroom feedback technique using Likert-scale ratings ranging from 1 (strongly disagree) to 5 (strongly agree). The direction of the rating scale for the evaluation items was opposite that of the rating scale for the emotions items, which many participants did not notice. Thus, some participants rated the techniques poorly although providing comments that they were excellent. Consequently, we asked the five evaluation items one more time at the conclusion of the study by e-mailing participants and asking them to complete the questions again on a Web-based survey. Most of the participants (71%) responded to our request, and we report the evaluation data using only these participants.

Results

A series of analyses did not find significant differences between groups regarding their age, sex, race, enrollment in a particular section of introductory psychology, or grade point average (ps > .05). There were also no preexisting group differences on self-reported knowledge of the lecture topic. One participant in the standard lecture condition entered late and thus did not provide data about emotions experienced before the lecture.

Student Participation and Review Question Performance

Using a one-way ANOVA and a Tukey post-hoc comparison, we found a significant effect of group on formal participation rates, F(2, 18) = 16.84, p < .001 (see Table 2). Although informal participation rates were comparable between groups, formal participation was highest in the clicker group, followed by the response card group, both of which were significantly higher than the hand-raising group. In contrast, the hand-raising group performed the best on the formal review questions (M = 98% correct, range = 88–100% across the seven review questions), whereas the clicker group did the poorest (M = 82% correct, range = 66–97%). The greatest differences between these two groups on the percentage correct occurred on questions that were apparently more difficult, with greater difficulty defined as a lower percentage of all students who got the question correct. On the most difficult question, there was a 22% difference in the percentage correct between these two groups (88% – 66%), whereas there was only a 3% difference on the easiest question (100% – 97%). These differences might relate to the honesty of feedback, which we discuss later.

Postlecture Quiz

Planned contrasts among the three review question groups (M = 57%, SD = 16%) and the standard lecture group (M = 57%, SD = 21%) found no significant differences on postlecture quiz scores (ps > .18). However, the clicker group’s performance on the formal review questions was closer to their performance on the postlecture quiz than the other groups (see Table 2), suggesting their answers to the review questions were a more honest (accurate) reflection of their actual learning.

Academic Emotions

Comparisons among groups on academic emotions experienced over time appear in Table 3. Entering group as a between-subject factor and time of measurement (before, during, after) as a within-subjects factor in an ANOVA did not result in any significant interactions. However, across time, enjoyment was lowest in the standard lecture condition, resulting in a
Table 2. Mean Emotion Ratings by Feedback Technique

<table>
<thead>
<tr>
<th>Emotion/Group</th>
<th>Before</th>
<th>During</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>1.85</td>
<td>1.74</td>
<td>1.40</td>
</tr>
<tr>
<td>Hand-raising</td>
<td>1.77</td>
<td>1.51</td>
<td>1.46</td>
</tr>
<tr>
<td>Response cards</td>
<td>2.00</td>
<td>1.51</td>
<td>1.37</td>
</tr>
<tr>
<td>Clickers</td>
<td>2.09</td>
<td>1.72</td>
<td>1.47</td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>1.99</td>
<td>1.77</td>
<td></td>
</tr>
<tr>
<td>Hand-raising</td>
<td>1.73</td>
<td>1.63</td>
<td></td>
</tr>
<tr>
<td>Response cards</td>
<td>1.73</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>Clickers</td>
<td>1.75</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>Boredom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>3.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand-raising</td>
<td>2.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response cards</td>
<td>2.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clickers</td>
<td>2.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>2.51</td>
<td>2.35</td>
<td>2.86</td>
</tr>
<tr>
<td>Hand-raising</td>
<td>2.92</td>
<td>2.60</td>
<td>3.21</td>
</tr>
<tr>
<td>Response cards</td>
<td>2.94</td>
<td>2.80</td>
<td>3.17</td>
</tr>
<tr>
<td>Clickers</td>
<td>2.82</td>
<td>2.89</td>
<td>3.09</td>
</tr>
<tr>
<td>Hope</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>3.21</td>
<td>3.21</td>
<td></td>
</tr>
<tr>
<td>Hand-raising</td>
<td>3.65</td>
<td>3.71</td>
<td></td>
</tr>
<tr>
<td>Response cards</td>
<td>3.54</td>
<td>3.58</td>
<td></td>
</tr>
<tr>
<td>Clickers</td>
<td>3.52</td>
<td>3.54</td>
<td></td>
</tr>
<tr>
<td>Hopeless</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>1.72</td>
<td>1.49</td>
<td>1.59</td>
</tr>
<tr>
<td>Hand-raising</td>
<td>1.59</td>
<td>1.22</td>
<td>1.40</td>
</tr>
<tr>
<td>Response cards</td>
<td>1.46</td>
<td>1.32</td>
<td>1.46</td>
</tr>
<tr>
<td>Clickers</td>
<td>1.45</td>
<td>1.32</td>
<td>1.41</td>
</tr>
<tr>
<td>Pride</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>2.62</td>
<td>2.88</td>
<td></td>
</tr>
<tr>
<td>Hand-raising</td>
<td>3.14</td>
<td>3.31</td>
<td></td>
</tr>
<tr>
<td>Response cards</td>
<td>3.20</td>
<td>3.23</td>
<td></td>
</tr>
<tr>
<td>Clickers</td>
<td>3.15</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>Shame</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>1.65</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>Hand-raising</td>
<td>1.48</td>
<td>1.67</td>
<td></td>
</tr>
<tr>
<td>Response cards</td>
<td>1.52</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>Clickers</td>
<td>1.54</td>
<td>1.61</td>
<td></td>
</tr>
</tbody>
</table>

Note. Means with different subscripts are significantly different from each other at $p < .05$ in a Tukey post-hoc comparison. Scale anchors are from 1 (strongly disagree) to 5 (strongly agree).

As shown in Table 1, the three groups that used active student responding methods had higher overall ratings of their respective classroom feedback technique than the standard lecture condition, $F(3, 95) = 3.44, p = .02$. However, in a Tukey post-hoc comparison, the only groups that significantly differed from one another on the overall subjective evaluation of the feedback technique were the standard lecture and hand-raising conditions ($p < .05$).

Discussion

Our findings suggest that certain classroom feedback techniques have moderately large effects on honesty of student feedback and participation rates and small effects on academic emotions. The particular techniques we employed, however, did not appear to have any significant effect on quiz performance after a single lecture. The few empirical studies on the use of clickers to improve student quiz scores have produced mixed results (Ewing, 2006; Kennedy & Cutts, 2005; Lee & Bainum, 2006). The postlecture quiz might also have been too difficult, or students might not have been sufficiently motivated to do well on a quiz that did not affect their course grade.

Compared to the other assessment techniques, the clickers slightly increased student enjoyment during the lecture. Pekrun et al. (2002) found that positive academic emotions are related to adaptive academic-related processes such as flexible strategy usage and sophisticated metacognitive monitoring. Thus, it might not be the experience of enjoyment (or any other emotion) that mediates the benefits of clickers, but rather the enhanced cognitive processing (attention) associated with it. Further research is needed to determine if the relatively small effect sizes related to emotional changes in a single setting have greater cumulative effects over time.

The most apparent advantage of using the clickers was the increased honesty of student feedback. In response to the presumably easy review questions, nearly everyone in each group provided the correct answer. As mentioned earlier, only 66% of the clicker group provided the correct answer to the most difficult question, whereas 88% of the hand-raising group provided the correct answer. Postlecture quiz scores suggested that the clicker group’s answers during the lecture more closely reflected how much they were actually
Table 3. Mean and Standard Deviation of Evaluation Items for Feedback Techniques

<table>
<thead>
<tr>
<th>Evaluation Item</th>
<th>Standard Lecture n = 21</th>
<th>Hand-Raising n = 28</th>
<th>Response Cards n = 27</th>
<th>Clickers n = 23</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1. It was easy to learn how to use —.</td>
<td>3.67&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.28</td>
<td>3.86&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.21</td>
</tr>
<tr>
<td>2. I feel I would do better on quizzes using —.</td>
<td>3.24&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.28</td>
<td>4.14&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.21</td>
</tr>
<tr>
<td>3. I would recommend — be used in future classes.</td>
<td>3.29&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.19</td>
<td>4.29&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.08</td>
</tr>
<tr>
<td>4. Other instructors should use —.</td>
<td>3.29&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.06</td>
<td>4.29&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.08</td>
</tr>
<tr>
<td>5. I like using —.</td>
<td>3.14&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.28</td>
<td>4.29&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.01</td>
</tr>
<tr>
<td>Overall M</td>
<td>3.32&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.06</td>
<td>4.17&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Note. Means with different subscripts are significantly different from each other across groups at p < .05 in a Tukey post-hoc comparison. Scale anchors are from 1 (strongly disagree) to 5 (strongly agree).

Learning, whereas those in the hand-raising group appeared to be influenced by social conformity. Indeed, this conformity was noticeable in the videotapes when students would hesitate to raise their hands (or response cards) until a sufficient number of other students did. Although hand-raising is a quick and easy way to assess student understanding, it also appears to convey to the instructor the illusion that students are “getting it” when they are not. Use of response cards, which are more anonymous than hand-raising but less anonymous than the clickers, also appears to be susceptible to social influence because of students’ hesitation until other students had responded. Over time, the cumulative effects of this illusion might have even more dramatic effects on students’ understanding of class material because instructors are unlikely to repeat or elaborate on content they believe students understand.

Another important finding was the difference in participation rates across groups. Although there were no differences for informal participation, only 76% of the hand-raising group responded when asked formal review questions. In contrast, the clickers and response cards increased formal participation to nearly 100%. Thus, another advantage of clickers and response cards is that they create an avenue for interaction with students who might be too shy to speak or even raise their hands.

Although our minilecture was representative of a realistic classroom lecture in terms of content for an introductory psychology class, the lecture was somewhat shorter than a full lecture. In addition, the data are from a single occasion, with the lecture not given by the participants’ regular introductory psychology teacher. Despite these limitations, the lecture provided during the study adequately approximated a realistic classroom lecture.

Based on our findings, we offer several recommendations. First, regardless of the classroom feedback technique, in-class review questions will likely increase student participation and reduce boredom. If technologically and financially feasible, a good choice for getting honest feedback, increased participation, and possibly greater student enjoyment is an audience response system with clickers. Although paper response cards are much less expensive and would be an acceptable second choice, they are also somewhat susceptible to the influence of social conformity.

References

presented at the 86th Annual Convention of the Western Psychological Association, Palm Springs, CA.

Notes

1. Jason M. Nelson is now at the Department of Psychology, University of Montana.
2. The authors thank Lindsey Leaf for assistance with the coding of the videotapes and William Addison for review of an earlier draft of the article.
3. Send correspondence to Jeffrey R. Stowell, Department of Psychology, Eastern Illinois University, Charleston, IL 61920; e-mail: jrstowell@eiu.edu.
Using Wireless Response Systems to Replicate Behavioral Research Findings in the Classroom

Anne M. Cleary
Colorado State University

College instructors are increasingly relying on wireless clicker systems as instructional tools in the classroom. Instructors commonly use clicker systems for such classroom activities as taking attendance, giving quizzes, and taking opinion polls. However, these systems are uniquely well suited for the teaching of psychology and other courses that emphasize behavioral research methods. Specifically, instructors can use the clicker system to engage students in an in-class replication of a known empirical phenomenon. This article describes 2 classroom demonstrations that reveal the usefulness of wireless clicker systems for replicating empirical phenomena in behavioral research.

Psychology courses that emphasize research present a challenge to instructors. Merely presenting data from a research study on a screen does not actively engage students in the means by which the researchers originally collected the data, and educators publishing in the higher education literature generally argue that actively engaging students during class time facilitates learning (e.g., McKeachie, 1999). Some institutions require laboratory components for research-oriented psychology courses (e.g., cognitive psychology, research methods, sensation and perception). The idea is similar to the principle underlying laboratory courses in biology and chemistry: Replication is one of the cornerstones of science, and active involvement in replications should increase students’ understanding of the scientific process, as well as their understanding of the particular phenomena in question. However, in psychology departments with particularly large enrollments, laboratory components are not always feasible, and instructors may often need to teach research-oriented courses using a large lecture-style format.

The advent of wireless remote keypad response systems (clickers) on college campuses (Beatty, 2004; “Q&A with Clicker Technology ‘Pioneers’,” 2005; cf. Wong, 2005, pp. 7–8) presents a potential means of actively engaging large classes in replications of known phenomena within psychology. These systems allow students to respond to questions or stimuli using clickers. A receiver unit routes each response to a computer, where software pools the data quickly, allowing the instructor to present the data on a screen for discussion (Beatty, 2004). The ability to instantaneously collect data from clickers means that instructors can use these systems to replicate known findings from human behavioral research in the classroom, even in classes containing hundreds of students. In fact, Langley, Cleary, and Kostic (2007) argued that clicker systems present a new technological means of carrying out experiments and collecting data for research studies within the behavioral sciences. Thus, although clicker systems are pedagogically useful for any type of course (as instructors can use them to take attendance, conduct opinion polls, and give quizzes), the systems are particularly well suited for the teaching of research-oriented courses within psychology because instructors can use them to replicate known research findings in the classroom. This article describes two in-class replications of cognitive psychology phenomena using a clicker system: the false memory effect (Roediger &...
McDermott, 1995) and the levels-of-processing effect (Craik & Tulving, 1975).

Demonstrations and Student Reactions

To replicate the false memory effect, I first read to the class five of the Roediger and McDermott (1995) lists (the “anger” list, the “fruit” list, the “spider” list, the “king” list, and the “high” list). Second, I presented the recognition test items on a screen. The test contained five old items, five critical new items, and five noncritical new items in a random order. Students used clickers to indicate “old” and “new” for each item. Third, I imported the class data into a Microsoft Excel spreadsheet, which I used to rapidly compute the mean proportion of “old” responses for each item type (old, critical, and new). Fourth, I created a graph of these mean proportions in Excel to show the class the false memory effect. Finally, I pasted the data into SPSS to run analyses and projected them on the screen. The differences shown in the graph were statistically significant. The comparison of primary interest, critical lures (M = .50, SD = .24) versus new lures (M = .17, SD = .20), replicated the false memory effect, t(34) = 7.65, p < .001.

Students’ feedback on the demonstration was generally positive. On a yes–no postdemonstration survey, 97% indicated that they found collecting data as a class and presenting it on the screen more compelling than a show of hands, and 95% recommended using the replication in future classes.

To replicate the levels-of-processing effect, I first instructed the students to use their keypads to answer yes–no questions about words appearing on the projection screen. Half of the questions asked whether a word contained capital letters; half asked whether a word fit into a given sentence. Then, I presented a randomly ordered recognition test containing six words from the letter condition, six from the sentence condition, and six new words. The students used their clickers to indicate whether a given test item was “old” or “new.” I used Excel to compute the mean proportions of “old” responses for each of the three test conditions (letter, sentence, and new). I showed a graph of the levels-of-processing effect to the class and pasted the data into SPSS for statistical analysis. The key finding was a greater proportion of items called “old” for previously presented words from the letter condition (M = .36, SD = .22), t(35) = 15.54, p < .001.

In a slightly more detailed yes–no postdemonstration survey, 97% of the students indicated that the replication was more compelling than a show of hands. Additionally, 97% recommended using the demonstration in future classes, 93% believed that the replication contributed to their understanding of the effect, and 90% thought that the replication would lead them to do better on a test question about the levels-of-processing effect.

Discussion

This article demonstrates the utility of clicker systems for in-class replications of known behavioral phenomena, even in large lecture-style classes. There are several advantages to using clickers over a show of hands. Whereas a show of hands reveals raw numbers, clicker systems allow for the rapid presentation of descriptive statistics (means and standard deviations) and inferential statistics through online analysis with statistical software in front of the class.

Use of computer software for presenting and analyzing the data allows for easy archiving of the data. I save the graphs of replications from my prior classes to show new classes how their data patterns compare to the data patterns from prior classes. I also post our results on WebCT for students. Also, clickers actively engage nearly all of the students in the class. Although students can refuse to respond with their clickers, they generally all participate. Finally, students indicated in their written feedback that presenting newly collected data on a screen for analysis was more compelling than a show of hands.

Although I did not directly measure learning outcomes, recent studies suggest that active participation in research may improve students’ understanding of psychological research (e.g., Rosell et al., 2005). In addition, other studies (e.g., Yoder & Hochevar, 2005) suggest that active learning techniques lead to better exam performance than such techniques as lecturing, providing video presentations, and assigning independent reading. Therefore, there is reason to suspect that actively engaging students in replications of known behavioral phenomena will lead to better understanding than merely presenting the data during a lecture or having students read about the study.
The clicker system does have its limitations. Although instructors and researchers may someday collect response-time data (as with the Stroop effect) using clickers, one cannot currently do so with existing systems. Also, one cannot yet use clickers to record typewritten verbal responses (as in tests of recall memory or word stem completion).

One can generally use clicker systems for in-class replications of effects involving yes–no, true–false, and multiple-choice responses, as well as responses involving ratings. Many clicker systems interface with Microsoft PowerPoint, enabling instructors to use pictures and auditory or video recordings as stimuli.

In summary, clicker systems permit actively involving students in the means by which researchers within psychology collect, analyze, and interpret their data. As college campuses continue to move toward clicker-equipped classrooms (Beatty, 2004; Wong, 2005), more instructors will have access to clicker systems for use in replications of behavioral phenomena.

References


Notes

1. The clicker system used in the demonstrations reported here was purchased with funds from National Science Foundation Grant 0349088 to Anne M. Cleary.
2. Send correspondence to Anne M. Cleary, Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO 80523–1876; e-mail: Anne.Cleary@colostate.edu.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
Efficacy of Personal Response Systems ("Clickers") in Large, Introductory Psychology Classes

Beth Morling, Meghan McAuliffe, Lawrence Cohen, and Thomas M. DiLorenzo
University of Delaware

Four sections of introductory psychology participated in a test of personal response systems (commonly called "clickers"). Two sections used clickers to answer multiple-choice quiz questions for extra credit; 2 sections did not. Even though we used clickers very minimally (mainly to administer quizzes and give immediate feedback in class), their use had a small, positive effect on exam scores. On anonymous course evaluations, students in 1 clicker section reported that regular attendance was more important, but otherwise, students in clicker sections (compared to traditional sections) did not report feeling significantly more engaged during class. We suggest that future researchers might combine clicker technology with other, established pedagogical techniques.

Personal hand-held responders, commonly called "clickers," are one of the latest trends in technology for teaching (Beatty, 2004; Duncan, 2005). Clickers are one potential tool for increasing interactive engagement with material, and courses that use interactive engagement show higher levels of concept learning (Hake, 1998). The instructor poses a question to the class (using Microsoft PowerPoint) and students respond with hand-held responders. After students have responded to the question, the instructor displays a histogram of the class's responses. Anecdotal evidence (based on faculty interest at teaching conferences) indicates a rising use of clickers on college campuses, but do clickers help students learn? Do they help students feel more engaged?

According to some researchers, students like clickers, and students also believe clickers make them feel more engaged. For example, on course evaluations, students at one university reported that clickers had more benefits than downsides (Draper & Brown, 2004). However, in most studies on clickers, researchers do not compare clicker groups to nonclicker comparison groups, so demand characteristics might explain the findings. When asked, "how useful do you think the handsets are?" (Draper & Brown, 2004) or if "clickers helped me learn" (Duncan, 2005), students might overestimate their perceptions of benefits of clickers, because such introspection is notoriously faulty (Nisbett & Wilson, 1977).

In this study, we compared four large sections of introductory psychology (two instructors taught two sections each). For each of the two instructors, one section used clickers and one section did not. The instructors used the clickers to administer multiple-choice questions on the reading; to display histograms of the question results; and, when relevant, to correct widespread misunderstandings. The primary dependent measures were exam scores (within each instructor, exams in the two sections contained identical items) and self-reports of interest and engagement collected via anonymous course evaluations at the end of the semester.

Method

Participants and Design

Participants were introductory psychology students (N = 1,290) at the University of Delaware enrolled in one of four large sections. Each section had approximately 320 students.

At our university, introductory psychology attracts mostly first-year students (80%), but upper class students also enroll. First-year students are, in essence, randomly assigned to introductory psychology sections. During summer orientation sessions, entering students make a list of courses they wish to take in the fall without regard to class time or professor. Then a computer...
Table 1. Responses (Ms) to Engagement Items Added to Anonymous Online Course Evaluations

<table>
<thead>
<tr>
<th>Item</th>
<th>Dr A Traditional</th>
<th>Dr A Clicker</th>
<th>t</th>
<th>Dr B Traditional</th>
<th>Dr B Clicker</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of students responding to evaluations</td>
<td>68%</td>
<td>70%</td>
<td></td>
<td>77%</td>
<td>67%</td>
<td></td>
</tr>
<tr>
<td>I paid attention in class, stayed engaged*</td>
<td>3.67</td>
<td>3.73</td>
<td>0.59</td>
<td>3.42</td>
<td>3.46</td>
<td>0.43</td>
</tr>
<tr>
<td>Regular attendance was important in this class*</td>
<td>3.10</td>
<td>3.33</td>
<td>1.74*</td>
<td>2.79</td>
<td>3.08</td>
<td>2.52*</td>
</tr>
<tr>
<td>I enjoyed coming to class*</td>
<td>2.67</td>
<td>2.78</td>
<td>0.97</td>
<td>3.10</td>
<td>3.26</td>
<td>1.61</td>
</tr>
<tr>
<td>How often did you read before class?</td>
<td>2.91</td>
<td>3.12</td>
<td>1.87*</td>
<td>2.57</td>
<td>2.60</td>
<td>0.37</td>
</tr>
<tr>
<td>How many classes did you miss?</td>
<td>3.71</td>
<td>3.94</td>
<td>1.49</td>
<td>3.51</td>
<td>3.77</td>
<td>1.70*</td>
</tr>
</tbody>
</table>

Note. The df for t values range from 432 to 486. All students (first-year and upper class students) are included in these Ms. *p < .10. **p < .01.

* Item was answered on a 5-point scale anchored by 1 (strongly disagree) and 5 (strongly agree).  
** Item scale: 1 = never; 2 = once or twice; 3 = about half the time; 4 = about 75% of the time; 5 = every day.  
† Item scale: 1 = I missed one or more times a week; 2 = I missed about once a week; 3 = I missed about once every two weeks; 4 = I missed between 3 and 5 times total; 5 = I missed 0 to 2 classes total.

program assigns them to a section of each course they requested. Although not purely random, such assignment of students to a section is more random than that of upper class students, who might have registered for specific sections based on their personal preferences. Because the nonsystematic assignment of first-year students provides good experimental control, we used only first-year students in our analyses of exam performance.

Clicker and Traditional Sections

Two of the sections (taught by Dr. A) met Monday, Wednesday, and Friday for 50 min at 9:05 a.m. and 10:10 a.m. The other two sections (taught by Dr. B) met Tuesday and Thursday for 75 min, at 9:30 a.m. and 12:30 p.m. Both professors taught using an interactive lecture style. Both professors taught the earlier section without clickers ("traditional" sections) and the later section with clickers. In clicker sections, at the beginning of class, the instructor posted five multiple-choice, fact-based questions, based on the day’s required reading. Students earned extra credit for answering these questions correctly. Later in the class period, if relevant, the instructor would briefly elaborate on a clicker question that most students had misunderstood. Other than this change, instructors taught the two sections identically.

Students in the nonclicker sections of the class were able to obtain the same number of extra credit points as those in the clicker sections. For extra credit, these students could participate in an extra research study or read a portion of a chapter of the textbook not assigned on the syllabus.

Materials

We used radio-frequency clickers manufactured by Classroom Performance System of eInstruction. Textbook publisher Allyn and Bacon also provided technical support. All sections used the same textbook (Gerrig & Zimbardo, 2005). The four sections covered similar material, but not always in the same topic order. Students in clicker sections purchased their responders bundled with their textbook.

Dependent Measures

Exams. Each instructor gave four multiple-choice exams (there was no comprehensive final exam). Within an instructor’s class, the two sections answered identical items (i.e., the same exam was presented at 9:05 and at 10:10). However, to reduce cheating, instructors distributed four different exam question orders in each exam session. In addition, exams in the earlier sections were labeled as Forms 1 through 4, and in the later sections were labeled as Forms 5 through 8. We analyzed the percentage of questions answered correctly.

Self-reports of engagement. Students completed anonymous, semester-end course evaluations online. In addition to the standard questions used for all courses, we wrote five questions specifically to measure engagement (see Table 1 for the items).
Table 2. Exam Ms and SDs for Clicker and Traditional Sections

<table>
<thead>
<tr>
<th></th>
<th>Exam 1</th>
<th></th>
<th>Exam 2</th>
<th></th>
<th>Exam 3</th>
<th></th>
<th>Exam 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Traditional</td>
<td>71.13</td>
<td>12.9</td>
<td>69.04</td>
<td>12.2</td>
<td>70.04</td>
<td>13.0</td>
<td>69.44</td>
<td>11.8</td>
</tr>
<tr>
<td>Clicker</td>
<td>72.69</td>
<td>11.7</td>
<td>69.62</td>
<td>11.6</td>
<td>70.82</td>
<td>14.0</td>
<td>72.43</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Note. Ns range from 560 to 574 for traditional sections and 476 to 482 for clicker sections (some students missed an exam). The values are based on first-year students only. Clicker sections scored significantly higher than traditional sections on Exam 1 and Exam 4 (see text).

Results

Exam Scores

We conducted a $2 \times 2 \times 4$ mixed MANOVA with instructor (Dr. A and Dr. B, between participants), clicker use (traditional vs. clicker, between participants) and exam (four exams, within participants) as the independent variables (we present Greenhouse Geisser values here). We eliminated students who eventually withdrew from the course. Our analysis showed significant main effects for clicker use, $F(1, 1027) = 6.21, p = .013$, partial $\eta^2 = .006$, and exam, $F(3, 3081) = 16.68, p < .01$, partial $\eta^2 = .016$, which were qualified by a Clicker Use $\times$ Exam interaction, $F(1, 3081) = 3.46, p = .02$, partial $\eta^2 = .003$ (see means in Table 2). We conducted four post-hoc contrasts comparing clicker to traditional sections for each exam separately, using the MS error term from the interaction and Bonferroni adjusted $p$ values. These contrasts were significant for Exam 1, $t(3081) = 3.01, p < .05$, Cohen’s $d = .13$, and Exam 4, $t(3081) = 5.78, p < .05$; Cohen’s $d = .26$. Thus, exam scores were higher for clicker sections than for traditional sections, but the main effect was driven by scores for Exams 1 and 4. The effect size for clicker use was very small.

An Exam $\times$ Instructor interaction, $F(3, 3081) = 26.29, p < .01$, partial $\eta^2 = .025$, showed that for Dr. A, Exam 1 scores were higher than scores for Exams 2, 3, and 4, whereas for Dr. B, Exam 1 and 3 scores were higher than those for Exams 2 and 4. No other main effects or interactions were significant.

To test whether the clicker effect was different for upper class students and first-year students, we conducted a mixed ANOVA with the full sample, using a $2$ (clicker section) $\times$ $2$ (professor) $\times$ $2$ (class: first-year or upper class student) $\times$ $4$ (exam) design. The effects for clickers did not significantly interact with the class variable, suggesting that the effectiveness of clickers did not depend on being a first-year or upper class student. In addition, the class variable did not interact with any other effects reported in the primary analysis (which analyzed first-year students only).

Self-Reports of Engagement

Because we received only summary output (over-all Ms and SDs, not individual responses) from the course evaluation items, we were unable to conduct a $2 \times 2$ (Professor $\times$ Clicker Use) ANOVA on the five engagement items. Consequently we did independent-groups $t$ tests for the five self-report items (see Table 1). In addition, because only summary output was available for these questions, our analysis included all levels of students (first-year and upper class students). Of 10 possible comparisons, 1 emerged significantly in favor of clickers (“Regular attendance was important in this class” for Dr. B, Cohen’s $d = .15$), three were marginally in favor of clickers (these three were related to class attendance and reading before class), and six showed no difference.

Discussion

Exam Performance

Our data suggest that using clickers to quiz students in class on material from their reading resulted in a small, positive effect on exam performance in large introductory psychology classes. We had a large sample size and took advantage of the near-random assignment of students to sections, reducing the possibility of selection effects. The outcome did not depend on which professor was teaching.

In our study, the instructors used clickers very minimally—to administer quizzes, publicly display the
results, and quickly correct any widespread misunderstandings. Thus, our study shows that using clickers modestly can result in a small gain in exam performance. The effect we found might be mediated by a number of mechanisms. Clickers may have provided opportunities for interactive engagement, an empirically supported method for promoting concept learning (Hake, 1998). Students may have been motivated by comparing their performance to their peers. Students may have simply benefited from being more prepared for class or from extra practice with the types of questions that might be on the exam. Future studies of clickers might be able to evaluate some of these specific mediators.

An alternative explanation for the improved exam performance in the clicker sections is that students in the earlier section might have shared exam questions with students in later sections. To investigate this explanation, we compared the first three exam scores from the Spring 2006 semester (i.e., the semester after the one in which we collected the present data) for Dr. B, who taught two traditional sections of introductory psychology in consecutive time periods, again with the same exams. Exam scores in the later section were higher for the first test, $t(477) = 2.88$, $p < .05$, Cohen’s $d = .27$, but not significantly different in the second and third tests, $t_s(476) = –1.17, –1.28$, ns, in contrast to the consistently higher scores for clicker sections on all exams in our study. This outcome provides some evidence against the “question-sharing” explanation for these results.

**Student Engagement**

Although Dr. B reported that students “got a kick out of them,” clickers had only marginal effects on self-reports of student engagement, attendance, and reading in this study—effects that may be attributable to Type I error. Students in clicker sections reported that class attendance was more important, but they did not report feeling significantly more engaged than students in traditional sections. Another possible indicator of engagement is attendance over the semester. Although professors did not take attendance, we inspected the participation rates for the clicker quizzes, and attendance neither increased nor decreased over the semester. It is possible that past studies indicating that students thought clickers help them learn might have been exaggerated by demand characteristics. Another explanation of the nearly null effect on engagement is that this was the first time Dr. A and Dr. B had used clickers; Duncan (2005) reported that students rate clickers higher when instructors are more experienced with them (perhaps because they write better conceptual questions or because they have mastered the technological details).

**Areas for Future Study**

Methodologically, our study was a strong test of the effectiveness of clickers by themselves, unconfounded by other pedagogical tools. However, future research should test the impact of combining clickers with other, well-established pedagogical methods.

One suggestion is to combine clickers with concept inventories; that is, standardized, multiple-choice questions that ask students about core concepts (see Foundation Coalition, n.d., for an example in engineering). Concept inventories include as distractors the most common wrong answers (e.g., “Which of the four correlations shows the strongest prediction: (a) $r = –.73$; (b) $r = .62$; (c) $r = .25$; (d) $r = .10$” adapted from Chew, 2004a). Chew (2004a, 2004b) has promoted the use of such conceptual questions in the psychology classroom.

Group discussion is another tool to combine with clickers. In our study, students worked on questions individually. Research might show that clickers lead to greater benefits when combined with cooperative learning. For example, Duncan (2005) suggested that students can enter their answers to a question both before and after small-group discussion.

Another teaching method that can be supported by clickers is just in time teaching (JiTT). Dr. A and Dr. B used clickers to supplement their lectures with a small amount of JiTT. JiTT is a method in which the instructor adapts the course lesson—sometimes right in the middle of class—to get at what students do not understand (e.g., Beekes, 2006). A future study might amplify the use of JiTT in combination with clicker questions.

**Practical Considerations**

To instructors who are considering using clickers in their classrooms, we recommend they consider how they will grade clicker performance. Because students in our study could earn extra credit points for answering correctly, we may have inadvertently induced an evaluative focus. Students in the clicker sections expressed anxiety about whether their responses were being correctly recorded, so the students may have seen the clickers as evaluation tools, not as learning
tools. In addition, several students in our study openly “cheated” by asking their neighbors in class what the right answer was, even though the instructors had not expressly allowed such discussion. We can think of two interpretations of the “cheating” we observed. If students simply copied each other, such cheating probably undermined the true learning value of clickers (Brothen & Wambach, 2001; Daniel, 2006). On the other hand, if asking one’s neighbors about the right answer introduced some cooperative learning to the students and made students think more deeply, it might have actually increased student engagement with the material. We encourage instructors to think about how they might handle potential student cheating on clicker questions.

Finally, we note that there are less expensive options that might be just as effective as clickers. For instance, a professor can quickly assess the understanding of a group of students by having them display color-coded index cards in response to a question (Kellum, Carr, & Dozier, 2001). Similarly, professors can foster group interaction and engagement with material using scratch-off quizzes (Epstein et al., 2002) without having to purchase, register, and maintain the clicker responders. Online quizzes, too, may give the same (or greater) benefit when used according to their best practices (Brothen, Daniel, & Finley, 2004).

Conclusion

In a design with near-random assignment, across two different professors, using clickers resulted in a small improvement in exam scores. Based on our results, instructors should not expect large improvements in exam performance if they use clickers mainly to administer and display reading quiz questions in introductory psychology classrooms. However, future research might establish that clickers have greater impact when they are combined with techniques such as cooperative learning or concept inventories.

Although students in one section reported that attendance was important, clickers did not otherwise improve reports of student engagement. We found little support for the engagement hypothesis, showing that clickers are not, by themselves, sufficient to increase subjective reports of engagement. Future studies could test whether adding clickers to well-documented teaching methods will increase subjective student engagement, as well as learning.

References


Notes

1. Portions of this research were presented at the National Institute for Teaching of Psychology, January 2006.
2. We are grateful to Allyn and Bacon for providing extensive technical and personal support for this project.

We specifically acknowledge the help of Sarah Ergen and Christine MacKrell. Lauren Reiss helped manage data, and Christopher Sanger provided significant technical support.

3. Send correspondence to Beth Morling, Department of Psychology, University of Delaware, Newark, DE 19716; e-mail: morling@udel.edu.
This study addressed how trait levels of classroom shyness can influence conformity when students answer opinion questions in different ways. We recruited 128 introductory psychology students to indicate their opinion on 50 controversial questions by raising their hand or anonymously pressing a button on a keypad (“clicker”). Compared to hand-raising, keypad responses had greater variability, suggesting that students were less likely to conform to the group’s opinion. Students who typically experience shame and anxiety in class did not conform any more than other students did, but they felt more uncomfortable raising their hands and indicated a stronger preference for using keypads when answering controversial questions.

Electronic student response systems (SRS) use keypads, or “clickers,” to efficiently record and display students’ answers to questions that can be used for concept checks, quizzes, opinion polls, and more (Caldwell, 2007). By pressing a button on a keypad, students anonymously send their response to a receiver attached to a computer that displays a histogram of the students’ responses. A growing body of literature on the use of SRS proclaims benefits of greater participation and increased emotional engagement during lecture (Stowell & Nelson, 2007), and possible benefits to student learning (Morling, McAuliffe, Cohen, & DiLorenzo, 2008; Poirier & Feldman, 2007).

In a comprehensive study using observational and self-report methods, Howard, Short, and Clark (1996) suggested that a number of factors contribute to the amount of participation in the classroom. The strongest predictor of greater classroom participation was being a nontraditional student (older than age 24), but the researchers also found that men participated more than women, and there was less participation in larger classes. Overall, the majority of students did not participate in class (66%), and of those who did, about half of them accounted for 89% of all students’ comments. Furthermore, among the most frequent reasons provided by students for not participating in class was the possibility that they would appear unintelligent to other students (Howard et al., 1996). Results from a qualitative study on students’ participation in discussions of controversial topics confirm that students are reluctant to openly express their diverse views because of perceived negative reactions from their peers, including being stereotyped and suffering confrontation (Lusk & Weinberg, 1994). Clearly, there is room for improvement in getting the “silent majority” to participate in class, especially when discussing controversial topics.

Shyness can also contribute to the lack of students’ classroom participation, and this could have consequences on their academic performance. For example, young children who were identified as shy by their teachers performed more poorly overall on a vocabulary test than did nonshy students. However, these differences disappeared when shy students took their written exams in a group setting that conferred greater anonymity when compared to taking a test alone in the presence of a teacher (Crozier & Hostettler, 2003).
One common touted advantage of keypads is that students who are too anxious or shy to verbalize during class might feel more comfortable using keypads to respond, by virtue of the keypad’s anonymity (Ewing, 2006). Although most students in one study agreed that the anonymity of using keypads made them less likely to be embarrassed in class (Magyar-Moe, Becker, Burek, McDougal, & McKell, 2008), ironically, another study found shy students, more than nonshy students, felt that using keypads made the classroom environment less personal and did not allow expression of opinions (Lee & Bainum, 2006). In another study, students who liked to hear other students’ opinions, although being hesitant to express their own opinion, did not rate the perceived helpfulness of the keypads any higher than nonreluctant students (Graham, Tripp, Seawright, & Joeckel, 2007).

In a previous study, when students used keypads to respond to in-class review questions, they were more likely to respond and, more important, were more honest in their responding than students in a hand-raising comparison group (Stowell & Nelson, 2007). Students in the hand-raising group appeared to be influenced by the number of other students raising their hands to answer. However, the researchers did not include a measure of shyness to learn if it was related to participation or conformity (Stowell & Nelson). Thus, to date, there are no published studies to support the idea that shy students prefer using SRS any more than other students do or that their responses differ from other less shy students’ responses, particularly when asked about their opinions on controversial topics.

This study addressed how trait levels of classroom-specific shyness might influence conformity when answering opinion questions. Based on the knowledge of social conformity in group settings (Asch, 1951), we anticipated that responses rendered by traditional hand-raising would be less variable than keypad responses, indicating greater conformity to the majority’s opinion. Second, we expected this difference to be greater for shy students, who might be more likely to conform (Mehrabian & Stefl, 1995). Third, we hypothesized that shy students would be more likely to favor the use of keypads over hand-raising because keypads provide greater anonymity. Finally, we anticipated that shy students would experience greater negative mood after responding publicly to controversial questions than students low in classroom shyness, consistent with findings that shyness and negative mood were positively correlated (Cowden, 2005).

Method

Participants

From our department’s introductory psychology research pool, we recruited 128 participants who participated in exchange for research participation credit to fulfill a course requirement. The majority of participants, 110 (86%), were first-year students or sophomores and 84 (66%) were women. The distribution of race or ethnic background was 93 (75%) White, 23 (18%) African American, 7 (4%) Latino/Hispanic, 4 (3%) other, and 1 (1%) Asian American. Depending on the day of participation, sample sizes were in the range of 7 to 10, 20, or 40 to 43 per group, resulting in class sizes nominally categorized as 10, 20, and 40 students.

Instruments

Classroom shyness. The Academic Emotions Questionnaire (AEQ; Pekrun, Goetz, Titz, & Perry, 2002) measures various emotions in academic settings such as a lecture or examination. The AEQ has no designated shyness subscale, but two of the subscale constructs, anxiety and shame, overlap significantly with shyness (Harder, Rockart, & Cutler, 1993; Henderson, 2002). We chose 14 items from the AEQ that ask about anxiety (e.g., “I get scared that I might say something wrong, so I’d rather not say anything”) and shame (e.g., “When I say anything in class I feel like I am making a fool of myself”) typically experienced in a regular classroom lecture. The sum of these two subscales represented classroom shyness, which had a Cronbach alpha of .92. In contrast to shyness, we also selected four AEQ items that measured classroom enjoyment (“I enjoy being in class”), $\alpha = .78$. Participants responded to all AEQ items on a Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Controversial statements. We used 50 controversial statements related to psychology taken from several texts in the Taking Sides Series (McGraw-Hill) and from our own design. Participants indicated the extent to which they agreed with the controversial statement on a scale from 1 (strongly agree) to 5 (strongly disagree). Sample statements include “People who are mentally ill should not be allowed to purchase guns,” “Sexual orientation is not a choice,” and “Exposure to media violence promotes aggressive behavior.” We used test bank software (ExamView 5.1) to randomize
the presentation order of the questions and then imported the questions into our SRS software program (TurningPoint™).

Mood. The Positive and Negative Affect Schedule (PANAS) contains 20 items that measure positive and negative mood (Watson, Clark, & Tellegen, 1988). For each item, participants were asked, “Indicate to what extent you feel this way right now, that is, at the present moment” on a Likert scale from 1 (very slightly or not at all) to 5 (extremely). We calculated positive and negative mood by summing the item responses for each 10-item subscale at each time point. The positive and negative subscales had alphas exceeding .83 and .76, respectively.

Procedure

On participants’ arrival in a classroom lecture hall, we randomly assigned them to a seat, with each seat having a designated keypad number and a group assignment that alternated every other seat. Thus, the groups were interspersed among each other in the seating arrangement. We videotaped the experiment so we could record individual hand-raising responses later without interrupting the experiment.

On the consent form, we stated that participants would be asked to use keypads and hand-raising to answer a total of 50 controversial questions. After providing informed consent, participants completed the mood and classroom emotion surveys. We provided instruction on using the keypads and hand-raising techniques, and then began with two practice questions to familiarize participants with the procedure. Next, the two groups took turns responding to each of the 50 controversial questions displayed on a large screen at the front of the classroom by using their keypads or by raising their hands at the appropriate time.

For each controversial statement, we waited 15 to 20 sec for the group using keypads to respond, and then asked the second group to indicate their opinion by raising their hand when we read the response choice that matched their opinion. We did not show the histogram of the keypad responses on the screen at any time, to prevent students’ keypad responses from influencing the hand-raising group’s responses. On every subsequent question, the two groups switched methods of responding (i.e., from keypads to hand-raising or vice versa). Thus, participants answered 25 questions in each manner to allow for within-subjects comparisons. After responding to the controversial opinion questions, participants completed the mood survey again, followed by a demographic survey.

Statistical Analyses

By definition, conformity occurs when group responses are similar to one another. Thus, greater deviation from a group’s central tendency of scores would indicate less conformity. We tested potential differences between the variability of opinions obtained from hand-raising with the variability obtained by keypad responding by calculating the median absolute deviation (MAD) for each participant’s response to a question. We calculated this measure of variability by finding the absolute value of the difference between the individual’s response and the group’s median response on each question. We chose the median over other measures of central tendency because it is less susceptible to extreme scores and we chose the absolute value because the direction in which the scores deviated from the median was not important in this study.

On all days of testing except the first day, we were able to track individuals’ keypad responses and perform repeated measures analyses, comparing their mean hand-raising MADs to their mean keypad MADs. Because of a technical problem on the first day of the study (n = 40), we could not match our keypad responses with hand-raising responses so we resorted to treating this day’s data as if we had collected them from two independent samples. Finally, we correlated participants’ trait levels of classroom emotions with other self-report measures, including demographic characteristics.

Results

On the first day of the study, we found significantly greater variability in the keypad responses (M ± SD; 0.75 ± 0.20) than the hand-raising responses (0.53 ± 0.18), F(1, 78) = 28.0, p < .001, η²p = .26. In a general linear model repeated measures analysis of data from the other days, we confirmed this finding in all class sizes that we tested (see Table 1). Group size, sex, race, age, and shyness did not interact significantly with the method of responding (ps > .09, η²p < .06, suggesting that the difference in variability between keypads and hand-raising methods was comparable across group size, demographic factors, and shyness. To determine if this difference occurred only because we asked a large number of controversial questions, we examined the
Table 1. Average Deviation From the Median for Keypad and Hand-Raising Items.

<table>
<thead>
<tr>
<th>Nominal Class Size</th>
<th>Keypad MADs</th>
<th>Hand MADs</th>
<th>F, p Value</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 students</td>
<td>0.68 ± 0.15</td>
<td>0.56 ± 0.19</td>
<td>$F(1, 24) = 5.62, p = .026$</td>
<td>0.19</td>
</tr>
<tr>
<td>20 students</td>
<td>0.81 ± 0.17</td>
<td>0.54 ± 0.14</td>
<td>$F(1, 19) = 46.24, p &lt; .001$</td>
<td>0.71</td>
</tr>
<tr>
<td>40 students</td>
<td>0.87 ± 0.25</td>
<td>0.73 ± 0.23</td>
<td>$F(1, 42) = 13.13, p = .001$</td>
<td>0.24</td>
</tr>
<tr>
<td>Overall</td>
<td>0.80 ± 0.23</td>
<td>0.64 ± 0.22</td>
<td>$F(1, 87) = 39.73, p &lt; .001$</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Note. MADs = median absolute deviations.

potential differences in variability after participants had answered just one question using each response method. The findings were similar, with greater variability in keypad responses than hand-raising, $F(1, 85) = 6.38, p = .01, \eta^2_p = .07$.

As expected, AEQ subscales of shame and anxiety were highly correlated, $r = .79, p < .001, N = 128$. The combined shame and anxiety score (shyness) was significantly associated with increased feelings of uncomfortableness when using hand-raising to answer controversial questions and a stronger preference for using keypads over hand-raising (see Table 2). Mean levels of negative mood after the experimental session did not change from baseline, and the interaction between shyness and time was not significant either, $p_s > .17$. Thus, the correlations between shyness and negative mood at the beginning and end of the session were similar ($r = .32, r = .26$, respectively, $p < .01$). In contrast, positive mood declined over time, $F(1, 126) = 14.4, p < .001, \eta^2_p = .10$, and the interaction between time and shyness was significant, $F(1, 126) = 8.2, p < .01, \eta^2_p = .06$. Thus, shyness correlated significantly with positive mood at the end ($r = .29, p < .01$), but not the beginning ($r = .14, ns$) of the session.

On average, women reported a higher level of shyness than men ($4.63 \pm 1.74$ vs. $3.80 \pm 1.17$), $F(1, 126) = 8.18, p < .01, \eta^2_p = .06$, and indicated a greater preference for using keypads to answer controversial questions ($3.12 \pm 0.84$ vs. $2.64 \pm 0.81$), $F(1, 126) = 9.74, p < .01, \eta^2_p = .05$. Increased age was correlated with decreased preference for keypads ($r = -.23, p < .01, N = 128$), but otherwise age and race were not significantly related to any of our dependent measures.

Discussion

Previous research demonstrated that students were likely to conform to the majority of the class when using hand-raising to respond to multiple-choice knowledge questions (Stowell & Nelson, 2007). In this study, we

Table 2. Correlations of Classroom Emotions, Keypad Survey Items, and Mood

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Classroom enjoyment</td>
<td>—</td>
<td>-.12</td>
<td>.26**</td>
<td>.02</td>
<td>-.01</td>
<td>-.08</td>
<td>.02</td>
</tr>
<tr>
<td>2. Classroom shyness</td>
<td>—</td>
<td>-.17</td>
<td>.33**</td>
<td>.21*</td>
<td>.25**</td>
<td>.26**</td>
<td>.25**</td>
</tr>
<tr>
<td>3. Amount of typical participation in class</td>
<td>—</td>
<td>-.13</td>
<td>.05</td>
<td>-.02</td>
<td>.18*</td>
<td>.19*</td>
<td></td>
</tr>
<tr>
<td>4. Raising my hand to express my opinion made me feel uncomfortable</td>
<td>—</td>
<td>.06</td>
<td>.20*</td>
<td>.29**</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Instructors should use keypads when asking questions of a controversial nature</td>
<td>—</td>
<td>.60**</td>
<td>.16</td>
<td>.18*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I preferred keypads to hand-raising</td>
<td>—</td>
<td>.21*</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Total negative mood</td>
<td>—</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Total positive mood</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Total negative mood and total positive mood represent the sum of the respective mood scores across Time 1 and Time 2.

*p < .05. **p < .01.
extended this finding to include measurable changes in behavior when students use keypads to answer opinion questions. Responding with keypads produced greater variability of group opinion, suggesting that students were less likely to conform to the group’s median opinion. This was true for groups ranging in size from 10 to 40, which counters a common criticism that keypads are only helpful in large classes. Indeed, in Bond’s (2005) meta-analysis of variations of Asch’s (1951) classic conformity study, he found that conformity increases dramatically with as few as three other influential sources, after which further increases in the number of sources have a diminished added effect. None of the theoretical models of conformity predict a point at which conformity would decrease (Bond), suggesting conformity might be even greater in very large classes (e.g., 200 students). In our study, class size did not significantly interact with the method of responding, but we did not collect data from classes with more than 50 students, mainly because of the logistics of videotaping and recording hand-raising responses to 50 questions.

One might argue that answering 50 controversial questions is not representative of what occurs in the classroom, and rightfully so. We chose a large number of questions to avoid making our findings dependent on any specific set of questions. However, our within-subjects analysis of just the first two questions revealed a significant difference in the variability of responding, suggesting that these differences are likely to occur when any number of controversial questions are asked in class.

We found that students who typically experience shame and anxiety in class felt more uncomfortable raising their hands and would prefer to use keypads to answer controversial questions. We also found that greater classroom shyness was associated with greater negative mood before and after the experiment. The consent form notified participants they would be asked to use hand-raising to answer controversial questions, and the anticipation factor might have accounted for the higher level of negative mood among shy students prior to the start of the experiment. The correlation between shyness and positive mood at the end of the experiment could be due to increased arousal, which overlaps with some of the positive mood items (e.g., “attentive” and “active”).

Similar to other personality traits, shyness is susceptible to situational influences. For example, shy adults did not differ from nonshy adults on linguistics measures when communicating in a computerized chat discussion without a webcam, but shy adults were less likely to self-disclose in the presence of a webcam (Brunet & Schmidt, 2008). In another study, an ambiguous opposite-sex verbal interaction affected shy students’ behavior more than giving a structured speech (Pilkonis, 1977). It is possible that other researchers have not found significant differences in keypad preferences of shy students because of their reliance on general measures of shyness that might fail to take into account the specific shyness behaviors that occur in the classroom. For example, general measures of shyness focus more on social interactions that would normally occur outside of the classroom, such as attending a party (Cheek & Buss, 1981), which might be considerably different than interactions within a classroom environment. Thus, general measures of shyness might not be as sensitive as the AEQ in detecting shyness differences in the classroom. As mentioned previously, Graham et al. (2007) did not find significant differences between reluctant and nonreluctant students’ ratings of how helpful the keypads were in promoting participation. Although Graham et al. asked a few questions related to classroom shyness, they did not ask students if they preferred using keypads over other traditional methods. More research is needed to determine if the potential link between shyness and keypad preferences is a reliable one.

Our measure of shyness correlated with student preferences for using keypads and increased negative emotion experienced in the classroom, but did not relate to the amount of conformity. Although one study revealed a moderate positive relationship between shyness and conformity (Mehrabian & Stefl, 1995), we found a nonsignificant correlation between shyness and the amount of variability in hand-raising responses ($r = .03$), suggesting that shyness and conformity in the classroom were not directly related. Thus, shy students’ preferences for keypads were more closely tied to their emotional reactions than their actual behavior, suggesting that shy and nonshy students are influenced equally by conformity, but shy students feel worse about their experience. These findings might be analogous to those of Pilkonis (1977), who found that shy and nonshy students performed equally well in a structured speaking task, but shy students reported greater negative mood after the task.

In conclusion, using keypads to answer controversial questions reduced conformity in the classroom, revealing a greater diversity of students’ opinions. The ability of SRS to make extreme opinions more visible to others in the classroom might lead to greater or more thoughtful subsequent discussion, but further research is needed to test this hypothesis. Even though women participants and shy students preferred using...
keypads more than their respective comparison groups, their levels of conformity were not significantly different, suggesting that conformity affected these groups of students equally. Future researchers could uncover the factors that account for individual differences in classroom conformity and how they relate to classroom participation.

References


Notes

1. The authors thank Tyson Holder and Lindsay Nash for assistance with the coding of the videotapes and Derek Bruff for reviewing an earlier draft of the article.

2. Send correspondence to Jeffrey R. Stowell, Department of Psychology, Eastern Illinois University, Charleston, IL 61920; e-mail: jrstowell@eiu.edu.
COMPUTERS IN TEACHING

Effective Student Use of Computerized Quizzes

Thomas Brothen
Cathrine Wambach
General College
University of Minnesota

Computerized quizzes are becoming more available to psychology students and instructors. We hypothesized that students’ ineffective use of such quizzes would predict poor course performance. Students in a personalized system of instruction life span human development course used 26 computerized, multiple-choice chapter quizzes to help them master the course textbook. Students who used a “prepare—gather feedback—restudy” strategy were more successful than students using quizzes to learn course material. We discuss these findings in the context of helping students become more effective learners.

In an effort to market their products, many textbook publishers’ Web sites offer interactive activities for students. Common to these sites are multiple-choice chapter quizzes that provide immediate feedback. Should instructors direct their students to them?

Students are likely to use computerized quizzes if they suspect these quizzes will be a quick way to learn material. For example, Marek, Griggs, and Christopher (1999) and Weiten, Dégua, Rehmke, and Sewell (1999) showed that students prefer to use study aids such as practice quizzes rather than reading to learn psychology (cf. Beyeler, 1998). Recent research suggests how computerized quizzes should be structured to avoid these problems.

Bangert-Drowns, Kulik, Kulik, and Morgan’s (1991) meta-analysis of research on the instructional effect of feedback revealed that type and timing of feedback were strongly related to learning. They concluded that students learned more if given the correct answer only after answering a question. Their analysis also suggested that corrective feedback is better than simply telling students they were right or wrong. By implication, quiz software should present the total score only when students finish answering the questions and then direct them to content areas in which they need more study. The software should foster a “prepare—gather feedback—restudy” strategy. However, if left to decide how to use the software, poorly prepared students can easily subvert this effective strategy by taking many quizzes hoping to improve or by pausing to look up answers during the quiz (a “quiz-to-learn” strategy).

Observations of students in our life span human development class suggested that they used one of these two strategies for taking our computerized quizzes. Some took the quizzes, recorded the feedback, and restudied before repeating (i.e., they used the prepare—gather feedback—restudy strategy). They seemed to be well prepared, taking quizzes quickly and completing chapters with maximum scores after only a few attempts. Other students sat with their books open looking up answers to questions, frequently asked for help, and seemed to be taking more quizzes than the other students did. They appeared to be using the quiz-to-learn strategy.

We hypothesized that students taking quizzes consistent with the prepare—gather feedback—restudy strategy would perform better than students taking quizzes as a way to learn the material. We tested this hypothesis in a course in which students had free access to computerized chapter quizzes that counted toward their grades.

Method

We taught a life span human development class through a modified form of the personalized system of instruction (Buskist, C. S. & DeGrandpre, 1991; Keller, 1968) in a networked microcomputer classroom (Brothen, 1992). Twenty-nine sophomores and juniors completed the course during the period of this study. In addition to participating in small group seminars and writing a term paper, students completed computerized chapter quizzes and unit exams. The 26 chapter quizzes (260 possible points) and 4 unit exams (200 possible points) accounted for 75% of the total course points.

Course management software (Brothen, 1995) that recorded scores and daily activity in individual log files also delivered the 10-item multiple-choice chapter quizzes that students completed on a self-paced schedule. Students could take chapter quizzes as many times as they liked in the classroom during class time or could work independently in campus computer labs. After the students finished a quiz they received their score and a list of explanations and page references for their incorrect answers. The items were randomly drawn from item pools of approximately 65 for each chapter. Students received 4 course points for 10 correct, 2 for 9 correct, 1 for 8 correct, and 0 for fewer than 8 correct.

The computerized unit exams were closed-book, 1-hr multiple-choice exams covering six or seven chapters. Students could take each exam only once. They saw their scores (but no feedback for wrong answers) when finished with each exam. Each of the unit exams consisted of 50 items from the
chapter quiz item banks. Students could keep track of all their scores with an online gradebook (Brothen, 1996).

The course syllabus and the instructors’ first-day course introduction emphasized that the best way to approach quizzes was to use them for feedback on effectiveness of studying and to complete them before the respective unit exams. We told students that the best way to use the quizzes was to study each chapter thoroughly, take the quiz without looking at their books, write down the feedback, restudy, and then try again. However, they could look for answers in their books and complete the quizzes at any time during the term. At term’s end we obtained students’ number of completed quarter credits and their grade point averages (GPAs) from the university records office (consistent with human subjects guidelines) and compiled data on their chapter quiz scores, when they took them, the amount of time they spent on each quiz, and their unit exam scores. We defined preparation as number of quiz attempts (fewer attempts indicating better preparation) and amount of time spent taking quizzes (less time indicating better preparation).

Results

The 29 students began the course with a mean of 62.17 (SD = 37.47) completed quarter credits and a mean cumulative GPA of 2.91 (SD = 0.40) on a 4.0 scale. Their mean grade in the course was 2.99 (SD = 0.63). Completed credits and GPA correlated with course grade (rs = .38 and .40, respectively, ps < .05) but not with any of the other variables discussed subsequently (all ps > .10).

Generally, students did better on the chapter quizzes and poorer on the unit exams. Overall, 83.29% of students’ highest chapter quiz scores were at the maximum score of 10. Only 4.38% of highest chapter quiz scores failed to reach the mastery criterion of 8 for students to obtain points. The four 50-point unit exam scores averaged somewhat lower (M1 = 36.38, SD1 = 6.35; M2 = 40.59, SD2 = 5.07; M3 = 37.90, SD3 = 4.98; and M4 = 37.86, SD4 = 5.52). Quiz scores and unit exam scores were related. The average correlation (Fisher’s r to z transformation applied to all averaged correlations) between all students’ mean scores on quizzes taken prior to each unit exam and their respective unit exam scores was significant, r(29) = .39, p < .05. The mean score on all quizzes taken by students correlated with their total points on unit exams, r(29) = .53, p < .01, and course grade, r(29) = .62, p < .01. Students spent an average of 8.67 min on each quiz (SD = 4.35) with individual averages ranging from 3.36 min to 18.32 min. Consistent with our hypothesis, the average time spent on quizzes correlated negatively with total points on unit exams, r(29) = −.30, p = .06.

The average student took 132.52 quizzes over the 26 chapters (SD = 65.17) with a mean score of 7.77 (SD = 1.17). Also consistent with our hypothesis, total number of quizzes taken was negatively correlated with mean quiz score, r(29) = −.36, p < .05. Additionally, taking many quizzes did not seem to lead to high unit exam scores. The average correlation between number of quizzes students took prior to each unit exam and their unit exam scores was nonsignificant, r(29) = .15, p > .10.

The relation between number of quizzes taken and quiz performance was not uniform. Correlations between total number of quizzes students took for each chapter and their mean quiz scores on that chapter were negative at the start of the term, became more negative about the time of the Unit 2 Exam, and then became less negative, reaching near 0 values by the last five chapters. A quadratic curve estimation model accounted for 62.8% of the variance for this pattern and was significant, F(1, 23) = 19.38, p < .001. The reversal was attributable to a subgroup of students who appeared to be using the quiz-to-learn strategy. Seven students who took more than the average number of quizzes early in the term took fewer quizzes, r(29) = −.46, p < .05; got lower scores on them, r(29) = −.53, p < .01; and reached lower maximum scores, r(29) = −.72, p < .01, as the term progressed. Their reduced effort and poorer performance on quizzes washed out the basic relation between number of quizzes and performance by term’s end.

Discussion

Our results indicated that spending more time taking quizzes and taking them more times was related to poorer exam performance. Looking up quiz answers and taking quizzes numerous times did not help exam performance even though the exams consisted of items identical to those in the quizzes. These data lead us to conclude that using computerized quizzes in lieu of reading and studying to learn course material is an ineffective study technique. Using such quizzes as practice devices to build fluency (cf. Johnson & Layng, 1992) might be effective, but our students almost always quit working on a chapter once they received the maximum score of 10. When students achieved a score of 10 by looking up most of the answers, the score was not necessarily indicative of mastery. Our course structure apparently encouraged some students to focus on simply getting 10s rather than attaining mastery of the material.

Thomas and Rohwer (1986) advocated teaching students a process called executive monitoring: Students appraise their need for further study, deploy strategies to meet those needs, and assess their learning progress. Computerized quizzes can be very helpful in this process. Reading, taking a quiz that gives informational feedback, restudying, and retaking the quiz fits Thomas and Rohwer’s model nicely. Although we encouraged our students to use the Thomas and Rohwer model the first day of class and in our course syllabus, some students regarded the quizzes as a task to complete and persisted even though their strategy did not result in high exam scores. We suspect this approach reflects ineffective study strategies that are long-standing, ingrained habits that will not be easily overcome when students can choose how to proceed. Computerized quizzes may even distract students from other, more traditional study strategies. Toward the end of the term when 7 of our students who behaved consistently with the quiz-to-learn strategy reduced their effort on quizzes, their performance worsened. Their performance decline suggests that they did not replace the quiz strategy with more effective methods, but simply reduced their overall effort.
In contrast, students in our general psychology course take proctored quizzes without books or notes. They get feedback when they finish and can repeat the quizzes. We find they improve both their knowledge of psychology and their test-taking ability under these conditions (Brothen & Wambach, 2000). For nonproctored quizzes, part of our job as instructors should be to convince students to use them for feedback in addition to, not instead of, traditional study methods. The data from this study can be useful in stimulating and backing up good advice to students. As for instructors, our advice is to scrutinize potential computerized quizzes for how they are administered and how they deliver feedback to students. If they are likely to encourage the quiz-to-learn strategy, we advise you to keep looking.

References


Note

Send correspondence to Thomas Brothen, 346 Appleby Hall, University of Minnesota, Minneapolis, MN 55455; e-mail: broth001@tc.umn.edu.
The Value of Time Limits on Internet Quizzes

Thomas Brothen and Cathrine Wambach
General College
University of Minnesota

This study evaluated 15-min time limits on 10-item multiple-choice quizzes delivered over the Internet. Students in a computer-assisted course in human development spent less time on quizzes and performed better on exams when they had time limits on their quizzes. We conclude that time limits are associated with better learning and exam performance because they reduce the opportunity to look up answers in lieu of learning the material.

There is wide availability of Internet quizzes offered to textbook adopters through publishers’ Web sites. There is also an array of courseware choices for instructors to create computerized Internet quizzes to meet their course goals (Edutools, 2002). For example, computerized quizzes facilitate administration of mastery learning models by encouraging students to prepare, take quizzes, restudy, and retake the quizzes until they achieve a desired score (Brothen & Wambach, 2000a; Kulik, Kulik, & Bangert-Drowns, 1990). Instructors can also deliver computerized quizzes in proctored or unproctored environments, set time limits, and make them available for practice or count them for course credit.

In a recent study (Brothen & Wambach, 2001), we explored student use of unproctored, untimed computerized quizzes that counted for course credit and demonstrated that rather than using the “prepare—gather feedback—re-study” strategy suggested by Thomas and Rohwer (1986), some students in our psychology of human development course used what we described as a “quiz-to-learn” strategy. Rather than studying and then using computerized chapter quizzes to gauge the effectiveness of their preparation and the feedback to guide further study, some students attempted quizzes without good preparation. Instead, they found answers to the questions in their textbooks. Our results were consistent with Bangert-Drowns, Kulik, Kulik, and Morgan’s (1991) research, which found that focusing on correct answers to quiz questions instead of studying the text is an ineffective learning strategy.

Our data (Brothen & Wambach, 2001) indicated that the quiz-to-learn strategy (i.e., students taking longer to complete unproctored quizzes) was associated with poorer performance on later, proctored unit exams. Our unproctored, untimed computerized quizzes may have tempted some students to short-circuit the study process. In addition, our using the quiz scores to partially determine the final grade may have added to students’ decisions to try to earn high scores as quickly as possible.

The access to quizzes by our human development students contrasts with the experience of students in our general psychology course. Our general psychology students take timed, proctored computerized quizzes in a computer classroom without books or notes, get feedback when they finish, and can repeat them twice to improve their scores. Students in general psychology improve both their knowledge of psychology and their level of preparation for quizzes as the semester progresses (Brothen & Wambach, 2000a). However, if instructors deliver computerized quizzes via the Internet instead of in a computerized classroom, how are they to encourage their students to use the quizzes as an opportunity for feedback?
One solution is to put time limits on quizzes. We reasoned that if students had less time to find answers in their textbooks, they would need to prepare better for the quizzes and thus learn the course material better. We used a quasi-experiment design to examine the hypothesis that time limits on quizzes result in better unit exam performance than having no time limits.

Method

Participants

The participants were students registered in two sections of a psychology of human development course. They were primarily female, traditional-age sophomores and juniors who enrolled in the course because it met a prerequisite for nursing and other health science professional programs. Fifty-eight students registered for two sections of the course. We excluded 12 students from the analysis because 8 canceled the course before completing any work and 4 others did not take one or more of the exams at its scheduled time. Forty-six students (25 in Section 1 and 21 in Section 2) took the chapter quizzes before each exam and completed all their exams on the designated days. We included only these 46 students (3 men and 43 women) in the study.

Procedure

Students read a life span development textbook (Feldman, 2000); we made quizzes for each chapter available to them on the Internet using WebCT (Edutools, 2002). WebCT randomly selects items for quizzes from large item pools. It allows instructors to set time limits, restrict quiz access to certain computers, and choose the type of feedback. The program saves all completed quizzes and keeps track of the amount of time a student spends on a quiz. We used the entire quiz item database supplied by the textbook publisher. Once loaded into the courseware, the quiz items could be edited for content.

We team-taught the two sections of the course using a modified form of the personalized system of instruction (PSI; Keller, 1968) in a networked microcomputer classroom (Brothen & Wambach, 2000b) one evening each week. We worked together managing the computerized PSI system for the two sections. We worked individually with students only on a term paper assignment to minimize how much our “teaching styles” might affect their textbook learning and quiz performance.

Students in both sections could take 10-item multiple-choice chapter quizzes on a self-paced schedule as many times as they liked at any computer with access to the Internet. The quiz software randomly picked 10 questions from the item pools of approximately 65 for each chapter. The quiz feedback showed all 10 items with students’ answers highlighted, whether each was correct, and the total number correct. Students did not see correct answers; rather, the instructions directed students to their textbook to restudy concepts for the questions they missed. Students received points toward their grade only for mastery scores of 8 or above on their best attempt for each of the chapter quizzes. They received 3 points for 10 correct, 2 for 9, and 1 for 8 correct. The 19 chapter quizzes (57 possible points) and four unit exams (160 possible points) accounted for 64% of the total course points.

Section 1 students had 15-min time limits for their quizzes whereas Section 2 students had no time limits. We set the limit at 15 min for two reasons. First, our most common request for accommodations for students with disabilities is for time and one half on tests (Brothen, Wambach, & Hansen, 2002). Because we advise students to spend 1 min on each item, 15-min limits are consistent with that accommodation. Second, our experience with WebCT revealed that students often inadvertently close a quiz or the browser and need time to log back in and finish their quizzes. We stressed for both sections that they should use the quizzes for feedback on effectiveness of studying and should complete the quizzes before the respective unit exams. Consistent with Thomas and Rohwer’s (1986) study method, we told students that the best way to use the quizzes was to study each chapter thoroughly, take the quiz without checking answers in their books, and complete the quizzes at any time during the term. Students could keep track of all their scores with the WebCT point-checking tool.

The four unit exams were closed book, 1-hr, 40-item computerized multiple-choice exams covering four to six chapters. Students took each exam in the classroom only once and saw only the total score when finished. The items for Unit Exams 1 through 4 were available on the WebCT website. Therefore, students could have seen all of them on the quizzes and simply memorized items rather than actually learning the material. To assess this alternative explanation, we rewrote the 40 Exam 4 items we used the previous semester. Our goal in rewriting the items was to create items that tested the same content and were at the same level of difficulty. We reasoned that if students were learning the material rather than simply memorizing items, they should perform as well on equivalent questions.

At term’s end we obtained students’ number of completed college credits, grade point averages (GPA), high school percentile ranks (HSR), and ACT scores from the university records office (consistent with our approved Institutional Review Board protocol). For chapter quizzes students took before the respective unit exams, we recorded scores, number of quizzes taken, and amount of time spent on each quiz. We also recorded the unit exam scores.

Results

The 46 students had completed a mean of 55.24 (SD = 28.04) semester credits with a mean cumulative GPA of 3.10 (SD = .57) on a 4-point scale. Their mean high school percentile rank was 80.24 (SD = 17.93) and their mean ACT score was 24.08 (SD = 3.64). The sections did not differ significantly on any of these measures of academic ability as indicated by t tests (all ps > .15).

Students with time limits on quizzes completed them faster. Time limited (TL) students took a mean of 4.33 min (SD = 1.49) to finish each quiz, and no time limit (NTL) students finished quizzes in a mean of 7.08 min (SD = 3.11). These means were significantly different, t(44) = 3.88, p < .001.
Time limits reduced the opportunity for TL students to use a quiz-to-learn strategy. If setting time limits enhances performance, the TL section (15-min limits) should have performed better on exams than the NTL section (no limits). Overall, TL students had statistically significant higher total exam points ($M = 140.92, SD = 9.20$) than NTL students ($M = 130.29, SD = 19.37$), $t(44) = 2.44, p < .05$.

To assess whether students were merely practicing for the exams by memorizing correct answers to specific items rather than learning the material, we compared scores on Exams 1 through 3 with scores on Exam 4 and amount of score change from Exams 1 through 3 to Exam 4 for the two sections by t tests. If students were learning items instead of the material, both sections should have done worse on Exam 4 than on Exams 1 through 3 because it contained new items not appearing on any of the quizzes. We found no overall differences in the averages on Exams 1 through 3 ($TL = 35.35, NTL = 33.20$) and Exam 4 ($TL = 34.59, NTL = 31.04$), or on score change ($p > .21$). The rewrite of Exam 4 did not result in significantly lower scores, suggesting that neither of the sections were dependent on item familiarity to do well on the exams.

Our data indicated that the differences between sections might have been larger if all NTL students had consistently used a quiz-to-learn strategy. TL students’ number of quizzes was uncorrelated with exam points whereas for the NTL section we found a significant positive correlation between number of quizzes taken and exam points, $r(21) = .60, p < .001$. Also, the NTL students’ total number of quizzes taken correlated negatively with mean quiz time, $r(21) = -.49, p < .001$. Thus sometimes NTL students’ quiz behavior was more similar to TL students than their NTL classmates. Three NTL students never evidenced the quiz-to-learn pattern of taking fewer quizzes and spending longer on them than TL students whereas on any one quiz approximately half the NTL students behaved this way.

Finally, because some students in the TL section could have been using a truncated version of the quiz-to-learn strategy just as some NTL students could have been eschewing it, we correlated mean time all students spent on quizzes with total points earned on exams. This correlation was high, $r(46) = -.86, p < .001$, and did not differ by section ($p > .10$). Overall, students who spent more time on quizzes (possibly incorporating a quiz-to-learn strategy) did poorer on exams. We tested the possibility that this high correlation was due to poorer students needing to spend more time on quizzes and still not performing as well as other students on exams by partial correlations that removed the effects of ACT score and HSR. The resulting correlations did not differ appreciably, $r_s(46) = -.83$ and -.82 respectively.

Discussion

As in our previous study (Brothen & Wambach, 2001) we found evidence of the quiz-to-learn strategy and of its negative association with learning and performance. In this quasi-experimental study, time limits on quizzes were associated with better performance on exams.

Our data suggest that time limits can improve learning by reducing the quiz-to-learn strategy. For example, one NTL student discovered that closing the browser resulted in the same quiz being delivered at her next login. She responded by printing the quiz, looking up the answers to the 10 questions and then entering them hours or days later (we removed her data from the time comparisons). She persisted in using this strategy even though she failed every unit exam. This example was probably an extreme case of the quiz-to-learn strategy; however, other NTL students showed consistent patterns of taking a couple of quizzes in average times with mediocre scores and then spending more than 15 min to get all the items correct. It was rare for NTL students to continue taking quizzes after getting a perfect score in this manner.

Our adoption of WebCT afforded us the opportunity to place time limits on quizzes. We suspect that time limits reduced the quiz-to-learn strategy not only by leaving less time for it but also by sensitizing students to the need to prepare well enough to be able to answer the questions on their own. We recommend that instructors set reasonable time limits on computerized quizzes designed to help students learn psychology better.

References


Note

Send correspondence or requests for further information to Thomas Brothen, 346 Appleby Hall, University of Minnesota, Minneapolis, MN 55455; e-mail: broth001@umn.edu.
The Value of Time Limits on Internet Quizzes
Thomas Brothen & Cathrine Wambach
Available online: 20 Nov 2009

To cite this article: Thomas Brothen & Cathrine Wambach (2004): The Value of Time Limits on Internet Quizzes, Teaching of Psychology, 31:1, 62-64
To link to this article: http://dx.doi.org/10.1207/s15328023top3101_12
Are Online Study Questions Beneficial?

Kristin Grimstad
Jamestown College

Mark Grabe
University of North Dakota

Brothen and Wambach (2001) found that under certain conditions students do not benefit from online resources. They awarded points toward the course grade for high performance on small sets of practice questions and concluded student strategies in completing these quizzes were not beneficial to course performance. We investigated student use of online questions when students received no credit for answering practice questions and found students who made voluntary use of online questions scored higher on course examinations. In contrast to situations in which students complete assigned tasks for credit, students who respond to questions as a voluntary study tactic may use question feedback quite differently and make decisions about how many questions to review using different criteria.

Brothen and Wambach’s (2001) study asked whether college instructors should direct students to the online sample quizzes made available by textbook companies. Their study was prompted by survey data (Marek, Griggs, & Christopher, 1999; Weiten, Deguara, Rehmke, & Sewell, 1999) indicating that students used sample test questions rather than other study resources because of their perception that reviewing questions would be a quick way to learn material. Brothen and Wambach argued that to the detriment of course performance students use a “quiz to learn” rather than a “prepare–gather feedback–restudy” strategy when using online questions (Brothen & Wambach, 2001, p. 292). Brothen and Wambach (2001) offered an online quiz system as part of a course using a modified form of the personalized system of instruction (Keller, 1968). A total of 260 of the 460 examination points available in the course were based on online practice quizzes. Only scores higher than 8 of 10 ques-
Participants

The 197 participants were students from a moderate-sized, Midwestern state university enrolled in a semester-long introductory psychology course. A total of 179 participants volunteered to complete the Nelson–Denny reading test (Brown, Fishco, & Hanna, 1993). Participants received extra credit as compensation for their time.

Materials and Instruments

Students completed three 50-item, multiple-choice examinations. Each examination covered four textbook chapters.

The instructor wrote some examination questions and selected some questions from the test bank accompanying the textbook. Practice questions available to students consisted of the odd-numbered questions from the textbook test bank. Students had access to approximately 100 multiple-choice questions for each chapter. The instructor selected course examination questions from the even-numbered items.

Participants connected to the online study system through an entry page requiring a valid user identification number and password. This entry page also served as the consent form for the study.

Each question appeared as an individual Web page. Following a correct response, a second Web page displayed the question and the simple message that the answer was correct. The computer program then randomly selected the next question from the pool available for the same chapter. Following an incorrect response, a second Web page displayed the question, indicated that the response was incorrect, and provided a page number or range of page numbers in the textbook containing relevant information. The following question was then selected from a set of questions covering the same section of the chapter.

A hypertext markup language (HTML) form inserted in each question and response page collected the student identification number, time (day, hour, minute, second), response category (question or answer page), unique question identifier, correct answer, and student answer. The summary of content understanding (Glenberg, Wilkinson, & Epstein, 1982; Pressley, Snyder, Levin, Murray, & Ghatala, 1987). Students often make inaccurate predictions of how well they will do on course examinations, and this inaccuracy is greater for students who perform poorly. Without an accurate self-assessment of preparedness, it is difficult to make effective strategic decisions about how long to study or to focus study time on areas of weakness. Exposure to practice questions can improve student prediction of general examination performance (Glenberg, Sanocki, Epstein, & Morris, 1987; Pressley et al., 1987), and linking failed questions with relevant information can improve test performance (Alessi, Anderson, & Goetz, 1979).

We believe online quizzes and online study activities are distinct and associated with different issues. For example, students are not always willing to make voluntary use of study resources (Rothkopf, 1988; Weiten et al., 1999), but will seldom ignore assignments that count toward their grade. We investigated voluntary use of online study resources and the relation of the use of these resources to course performance.

Method

Participants

The 197 participants were students from a moderate-sized, Midwestern state university enrolled in a semester-long introductory psychology course. A total of 179 participants volunteered to complete the Nelson–Denny reading test (Brown, Fishco, & Hanna, 1993). Participants received extra credit as compensation for their time.
Group means are the adjusted means associated with the ANCOVAs (see Meyer & Well, 2003, p. 419).

Students reacted to question feedback in different ways. When informed answers were incorrect, some students consistently moved immediately to the next question, whereas other students delayed before considering another question. We defined two groups, the short delay group and the long delay group, based on whether participant median delays following incorrect responses were above or below the median delay score for all users. The median delays on the three examinations were 14.5 sec, 14.5 sec, and 11.5 sec, respectively. Separate ANOVAs demonstrated no significant group difference in examination performance.

The number of questions attempted in preparation for each of the three examinations and performance on the study questions were either statistically unrelated or positively correlated. Correlations between the number of study questions attempted and performance on the practice questions available during each course segment were .18 (n = 61), .15 (n = 73), and .35 (n = 60), respectively. Only the correlation associated with preparation for the third examination was significant (p < .03).

ANOVA procedures compared the time students spent answering questions correctly and incorrectly. The mean time in seconds answered questions correctly and incorrectly for each of the three examinations were 14.4 (SD = 7.0) versus 20.8 (SD = 12.6), 15.3 (SD = 5.4) versus 20.2 (SD = 8.0), and 13.3 (SD = 7.2) versus 17.7 (SD = 13.5), respectively. On all three examinations, students took longer to respond to the questions they answered incorrectly: first examination, F(1, 60) = 41.7, p < .001, \( \omega^2 = .40 \); second examination, F(1, 72) = 65.2, p < .001, \( \omega^2 = .47 \); and third examination, F(1, 59) = 13.7, p < .001, \( \omega^2 = .17 \).

Discussion

Convenient Internet access could increase the study resources available to college students. Brothen and Wambach (2001) observed that some textbook companies already make study questions available and sought to determine whether students would use these questions in an effective manner. Their research led to the conclusion that online graded quizzes represent an ineffective study technique. However, because they awarded points for quiz performance, their grading structure may have influenced how students reacted to the questions. Their research may not provide insight into how students would use online study quizzes as a voluntary study tool.

Based on the minimum criterion of 50 questions per examination, at least 30% of students in our study took advantage of voluntary access to study questions for each of three examinations. Moreover, students who took advantage of access to study questions performed consistently better than nonusers on course examinations even after accounting for differences in reading aptitude.

Our study suggests that students reacted differently to mastery quizzes and voluntary study questions. Brothen and Wambach (2001) reported a negative correlation between the number of questions attempted and performance on the quiz questions. In contrast, our data showed either no significant relation or a positive relation. When the motivation is to earn points contributing to the course grade, knowledgeable students are likely to reach criterion quickly. Less knowledgeable students are forced to keep repeating practice tests, creating the negative correlation between quiz performance and quiz questions attempted. In contrast, when students have access to hundreds of questions intended only for practice and feedback, they may continue to examine items as long as they perceive this behavior is helping them learn.

It is possible Brothen and Wambach (2001) were correct in suggesting that students who knew less used their books to find answers to get points from the quizzes. Because students used this strategy to serve an immediate goal, students who frequently resorted to this strategy did not necessarily focus on improving their understanding and consequently performed poorly on unit examinations. Students may simply look up the answers to unfamiliar questions for the initial quiz associated with each unit to avoid having to complete additional quizzes. In contrast, when using questions as a voluntary study tactic, there is little incentive to look up information related to practice questions to answer the questions. It seems likely that students would respond quickly when correct answers are obvious and to delay when struggling unsuccessfully to identify a correct response. It is not surprising our study demonstrated that the time spent to answer questions correctly was less than the time spent to answer questions incorrectly. However, in contrast to Brothen and Wambach, our interpretation of the additional time spent by students in providing incorrect responses would not imply that students were engaged in maladaptive behavior (i.e., looking up information to avoid having to re-take quizzes).

The data collection technique in our study allowed the differentiation of time spent in answering each question from the time spent after receiving information about response accuracy. We contrasted those participants who took more time after incorrect responses with those participants who moved on quickly to determine if those who delayed performed better on examinations. The group of students who spent more time after incorrect responses did not perform better on course examinations. However, there is no way to know from these data if those who delayed longer were systematically looking up answers to questions they missed or engaging in any form of active remediation.

Students who answered at least 50 practice questions performed better on course examinations. However, students who consistently spent more time after learning an answer was incorrect, presumably to engage in question-specific remediation, were at no observable advantage. This study thus supports the position that voluntary use of online study questions is associated with higher examination scores, but does not provide insight into the specific mechanism that might explain this advantage. Additional research to better understand how students make voluntary use of online practice questions and how strategies of use relate to performance is needed.

References


**Note**

Send correspondence to Kristin Grimstad, Computer Science Department, Jamestown College, Jamestown, ND 58045; e-mail: grimstad@jc.edu.
Are Online Study Questions Beneficial?
Kristin Grimstad & Mark Grabe
Available online: 20 Nov 2009

To cite this article: Kristin Grimstad & Mark Grabe (2004): Are Online Study Questions Beneficial?, Teaching of Psychology, 31:2, 143-146
To link to this article: http://dx.doi.org/10.1207/s15328023top3102_8

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions
This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
of some of our academic guide maps in teaching techniques of college learning, general psychology, and memory and cognition courses, and students’ feedback suggests that these types of guide maps are useful in organizing information and making connections between topic areas.

References


Notes

2. Send correspondence to Michael A. Motes, Department of Psychology, TCU Box 298920, Texas Christian University, Fort Worth, TX 76129; e-mail: m.a.motes@tcu.edu.

Forbidden Words: A Strategy for Studying Psychology

Michelle M. Merwin
University of Tennessee at Martin

I adapted the board game Taboo© for 3 psychology classes: introductory, abnormal, and history. Teammates guessed terms described by a clue-giver; however, the clue-giver had to avoid using 5 “forbidden words” listed on playing cards. The forbidden words were closely related to a term and would have been obvious clues to use. Using terms from text chapters, students generated forbidden words and later played the game. Evaluation indicated that students believed that creating the clues helped them learn the concepts. Students indicated that playing the game would serve as a good test review and that they would like to play again.

Describe Wilhelm Wundt without using the phrases “introspection,” “structuralism,” “Leipzig, Germany,” “father of psychology,” and “laboratory.” Without restrictions, one could simply describe him as the father of psychology or the person who started the first laboratory in Germany. Forbidding the use of obvious words in a guessing game is the premise of a classroom activity developed that capitalizes on basic learning tools: expressing ideas in one’s own words, thinking of examples, applying the word to new contexts, repeatedly studying the words, actively rehearsing and thinking about the words, and testing one’s knowledge (Myers, 2001).

I adapted the board game Taboo© to psychology material (using terms and concepts found at the end of text chapters) and named the resulting game Forbidden Words (FWs). The object of the game, in which two teams compete, is for team players to correctly guess as many words as possible in 1 min. Each player in turn describes words drawn from a deck to teammates. A player describes the word at the top of the card (hereafter called term) to teammates who try to guess it. Five FWs, which cannot be used in the clues, appear beneath the term.

Note the integral relation between the term and the forbidden words. For the term acetylcholine, the FWs might be brain, Alzheimer’s disease, neurotransmitter, muscle, and memory. The clue-giver describes acetylcholine without using any of the FWs (e.g., “a chemical messenger that helps people remember”). The clue-giver gives as many clues as necessary, but cannot use any part of the term or FWs. Avoiding the FWs is what makes the game challenging and fun. When the team guesses correctly, the clue-giver continues with subsequent cards for the remainder of the minute. The team earns points for correct responses and loses points for using an FW or skipping cards (1 point for each correctly guessed term, 1 point deduction for skipped cards or when a player uses an FW). An opposing team member looks over the clue-giver’s shoulder to ensure that no FWs are used.

Initially I used FWs in introductory psychology. I later expanded its use to abnormal psychology and history of psychology courses. I developed FWs in three phases. After introducing the chapter, I assigned terms to students and they generated the FWs. Generating FWs early on forces stu-
Students to preview the chapter and engages them in creating the game. Furthermore, it gives the instructor time to make the cards. Second, from the students’ work I selected quality FWs. Third, students played the game in one class session prior to a test. An abnormal psychology class completed an anonymous evaluation of this activity.

Method

Timeline

I use FWs as review before exams (class size = 35 to 65). It does not replace primary instruction. About 2 weeks before an exam, I review a handout and assign four terms to each student, due at the next class. For each term, students generate five FWs and briefly explain why they chose them. In larger classes more than one student may generate FWs for a term; overlap provides the instructor variety when selecting quality FWs. Note that students may be generating FWs for terms not yet covered in lecture, but by the time the cards are ready for play, the students will be familiar with the terms (either through lecture or reading). Most students are able to glean information from the text to make adequate FWs.

Selecting FWs

To create quality FWs, students need to extract relevant information about the term from their text. For example, in creating the cards, it may be helpful for student and instructor to ask, “What are the essential elements related to acetylcholine?” and “What words would a player find most useful in describing it to another player?” Quality FWs are ones that most players would like to use and are integrally related to the term. For example, based on my text, students should note that acetylcholine is a neurotransmitter that plays a role in memory, muscle contraction, and Alzheimer’s disease. Thus, the FWs mentioned earlier are of high quality. FWs should include some understanding of important functions, description, purpose, or synonyms. After examining the student-generated words, I selected five exemplary FWs and made the cards. I did not limit the FWs to the use of single words, allowing phrases such as “father of psychology.”

For a class of 65 students I made cards for six groups, each comprised of two teams having 4 to 6 players on each team. For each group, I mixed and divided the cards, giving each team half. The students played the game, and a class evaluation (N = 50) showed that half (51%) believed that playing FWs helped them learn the concepts and few indicated that they would use FWs if it was on library reserve (26%).

Discussion

FWs is an adaptable activity. Instructors can use it to review a range of material (single chapter, units, comprehensive final, or senior exit exam) and any psychology subject. Furthermore, FWs is adaptable to class and instructor preparation time. Instructors can generate a single set of cards to use repeatedly. Alternatively, if time permits, having students generate the FWs helps them preview the material and heightens their involvement. Developing the FWs requires that students think about the ideas and generate optimal words to describe them. The task encourages students to boil down the concepts into five essential words or phrases. Student selection of FWs varied greatly; some students required additional instruction in developing quality FWs. Some students generated FWs that were well conceptualized (for the term cones, FWs were rods, color, eye, retina, and daylight). Others placed too much emphasis on the textbook description, resulting in FWs that were too concrete (for the term signal detection theory, FWs were weak signals, life-or-death consequences, whimpering baby, detection, and experience).

FWs serves as a good review. It forces students to express the terms in their own words, actively study the material, and implicitly test their knowledge. Generating synonyms probably deepens the level of processing and semantic encoding (Craik & Tulving, 1975). Even when students are not directly engaged in play (i.e., they are waiting while the other team gives and receives clues), they can play along and test their knowledge. Playing the game with other students is fun and humorous, an important element in learning (Kaplan & Pascoe, 1977; Ziv, 1988). Additionally, the game offers opportunity for clearing up confusing topics. The instructor can listen to clues and guesses and clarify topics when necessary.

The evaluation results do not adequately describe the level of enthusiasm expressed by the students while playing the game. Although students took the activity seriously, there was much laughter and excitement while playing. However, the ill-prepared and unmotivated do not enjoy FWs. To play well, students must be familiar with the material and expend effort. It is beneficial if they review the terms in preparation for the game. The peer pressure of performing often motivates them to study. Whether the student is giving or receiving clues, FWs helps students to think differently about the material.

References

Not Another Group Project: Why Good Teachers Should Care About Bad Group Experiences

Krista D. Forrest
Richard L. Miller
University of Nebraska at Kearney

This article presents theoretical and empirical support for the idea that past group experiences can distort students’ perceptions of current group work. We assigned students to positive and negative current group experiences. As expected, students in the positive current group condition were more satisfied with their group’s process. However, current group experience had no effect on performance satisfaction or commitment to future group work. In comparison, past group experiences significantly influenced measures of process and performance satisfaction as well as commitment to future group work. Recommendations for reducing the effects of negative group experiences include having students talk about their past group experiences and providing students with strategies for creating effective group processes.

Group projects benefit students by facilitating critical thinking (Cooper, 1995), encouraging better processing and evaluating research experiences (Gibson, Kahn, & Mathie, 1996). As a result, more teachers appear to be combining collaborative learning with traditional teaching (Meyers, 1997). Between 1974 and 2002, Teaching of Psychology included 80 articles describing different collaborative strategies. Examples of such collaborative learning included group research projects, role-playing exercises, discussion groups, and debates (Meyers, 1997).

Research suggests, however, that students do not always enjoy these experiences (Phipps, Phipps, Kask, & Higgins, 2001; Wheelan & Lisk, 2000). For example, Forrest, Kershaw, and Bott (1998) asked college students to describe their most positive and negative group experiences. Events associated with academic work groups comprised 54% of the negative group experiences and only 26% of the positive group experiences. The major, unique feature of these negative academic work groups was the perception of social loafing.

Williams and Karau (1991) examined how an expectation for social loafing often leads a group member to compensate or work harder in a group than he or she would have worked alone. In some situations, the effort expended when working in a group is consistent with a “preexisting effort script” (Karau & Williams, 1993, p. 685). This script involves an individual’s expectations for how much he or she will need to contribute for the group to be successful and is based on past group experiences. In brief, students may develop negative attitudes toward group work because they have had to compensate for loafing members to obtain acceptable outcomes on group assignments. As a result, they may conclude that they will need to contribute more than their fair share in future similar projects. Accordingly, teachers need to be more cognizant of the types of group experiences that students have had, along with the effects of these experiences on satisfaction with current and future groups.

The goal of this study was to examine the extent to which students’ past as compared to current group experiences influenced ratings of group process, performance, and commitment to future group work. To investigate whether past group experiences were more influential on these ratings than were current group experiences, we assigned some participants to a positive current group experience condition and others to a negative current group experience condition. We hypothesized that students whose past group experiences were negative would be less satisfied with their current group’s process and performance even if the current group experience was positive, and in turn, be less willing to work on another group activity.

Method

Design and Measures

Our design was a 2 (past group experience: positive, negative) × 2 (current group experience: positive, negative) factorial on measures of group process, performance, and commitment to future group work.

Past group experience. To evaluate past group experience, we asked participants “In general, how do you rate your past experiences working in groups?” using a 7-point scale ranging from 1 (negative) to 7 (positive). We used a median split to categorize participants as having either positive or negative past group experiences.

Current group experience. Based on previous research (Forrest et al., 1998), our goal was to systematically create group experiences that included or did not include social loafing. In the positive current group conditions, the leader expressed interest in the task by stating that the task was interesting, maintaining eye contact with other group members who were speaking, and taking copious notes. In the negative current group condition, the leader expressed disinterest in the task by stating that the task was boring, refusing to maintain eye contact with other group members when they spoke, and taking limited notes. Regardless of the condition, group leaders were not allowed to contribute to the task solution itself.

Dependent measures. The statement “I was pleased with the way the group discussed the problem” served as the measure of group process. The statement “I was pleased with
Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned

David B. Daniel  
University of Maine at Farmington

John Broida  
University of Southern Maine

This study examined the utility of Web-based quizzing. We assigned 3 classes to a no-quiz, in-class quiz, or Web-based quiz condition. Midsemester results demonstrated a positive effect for in-class quizzing but not Web-based quizzing. After several adjustments in quiz presentation and duration, the Web-based group increased exam performance to a level equivalent to the in-class quiz group for the second half of the semester. These results illustrate that online quizzing can be as effective as in-class quizzing, but only under specific conditions.

A number of studies have provided evidence that routine quizzing increases student performance on exams (e.g., Connor-Greene, 2000; Grover, Becker, & Davis, 1989; Taraban, Maki, & Rynearson, 1999). However, quizzing often consumes valuable class time and requires that someone grade and record student performance. The increased effort required from the instructor diminishes the practical utility of quizzing as class size increases. Thus, quizzes are not often used in courses with large enrollment where the positive impact in exam performance may be most appreciated.

Alternatively, Web-based quizzing outside of class may be a useful tool for incorporating the benefits of quizzing without sacrificing valuable class time (Brothen & Wambach, 2001). An additional benefit of this system is that the quizzes can be automatically scored and recorded for the instructor, reducing the burden placed on the instructor as class size increases.

Many faculty, however, express concerns that computer-based teaching tools may not be as effective as traditional teaching strategies (Brewster, 1996). We examined the utility of Web-based quizzing and its effects on multiple choice exam scores in a moderately sized psychology course.

Method

Participants

One hundred and twenty-five students enrolled in three sections of Child and Adolescent Development at a public liberal arts university in New England participated. We assigned each of the three sections to a no-quiz (n = 44), an in-class quiz (n = 42), or a Web-based quiz (n = 39) condition.

Procedure

The in-class quiz group received 16 weekly chapter-based quizzes in the first 15 min of class. Students in the Web-based quiz group received the same quizzes available for self-administration on the Web 24 hr preceding class, also allowing 15 min for completion. The outcome measures were four exams. For the first half of the semester, we presented identical quizzes to both quiz groups, and all three groups completed the same exams. In addition, the same lecturer taught all classes using the same notes, with the lectures lasting approximately the same amount of time in each class (60 min twice weekly).

Results 1

As seen in Table 1, although in-class quizzing seemed to demonstrate a positive impact on exam scores at midsemester, a cursory view of exam performance did not yield an obvious impact of Web-based quizzing on exam scores when compared to the no-quiz group. We computed the mean of Exams 1 and 2 for each group for the analysis. A one-way ANOVA revealed significant effects for quiz condition $F(2, 122) = 46.69, p < .01, \eta^2 = .43$. Bonferroni post hoc tests indicated significant differences between the in-class quiz group and both the Web-based quiz group, $t(79) = 7.46, p < .001$, and the no-quiz group, $t(84) = 9.38, p < .001$. There were no significant differences between Web-based quizzing and the no-quiz groups, $t(81) = 1.64, p > .05$. Consequently, we decided to pursue possible explanations to account for the lack of impact for Web-based quizzing on exam scores.

A Midsemester Correction

On the second exam, students in both quiz groups anonymously described common methods to “cheat” on the quizzes. Students in the Web-based group reported a number of strategies for cheating; most prominently reported were printing and sharing of quizzes, looking up answers in the book during the quiz, using an online glossary opened in a window adjacent to the quiz, and working in groups.
As a result of this information, we changed the Web-based quiz parameters. In particular, we added an additional pool of 100 questions from the supplied testbank to the publisher-provided 10 question quizzes, enabling each student to receive a different random selection of questions on each quiz. Each student’s quiz, therefore, consisted of different questions. In addition, we removed the glossary from the Web site and reduced the time allowed for each 10-item quiz from 15 to 7 min. With these changes, the students resumed the quizzing schedule outlined in Method 1.

Results 2

We computed the mean of Exams 3 and 4 for each group and performed a one-way ANOVA on the means. Results of this analysis revealed significant effects for quiz condition, \( F(2, 122) = 81.70, p < .01, \eta^2 = .57 \). Bonferroni post hoc tests indicated significant differences between the no-quiz group and both the Web-based quiz group, \( t(81) = 11.47, p < .001 \), and the in-class quiz group, \( t(84) = 10.45, p < .001 \). Unlike the results in the first half of the semester, there were no longer significant differences between Web-based quizzing and the in-class quiz groups, \( t(79) = .54, p > .05 \). In sum, the Web-based quiz group increased exam performance to a level equivalent to the in-class quiz group and significantly higher than that of the no-quiz group (see Table 1).

Discussion

Although quizzes are generally effective, Web-based quizzes do not always positively affect exam performance as compared to in-class quizzes (see also Brothen & Wambach, 2001). Students in the Web-based quiz group used strategies to optimize their quiz performance without mastering the text. Efforts to discourage these efforts by randomly assigning questions from a larger test bank and decreasing the amount of time allowed for the quiz were effective.

Unfortunately, much of the Web content currently available from publishers in various formats is of the type that does not discourage such activities by students: Every student receives the same questions in the identical order with no time limits imposed. Furthermore, not all Web-based platforms allow for the corrections evaluated in this study (see Brooks, 2001). We argue that these, and possibly other, adaptations of publisher-provided content are necessary to fully obtain benefits similar to in-class quizzing with Web-based products.

Notes

Note

Please send correspondence to David B. Daniel, Department of Psychology, University of Maine at Farmington, 234 Main Street, Farmington, ME 04938, e-mail: dbdaniel@maine.edu.

References

Teaching of Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/htop20

Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned
David B. Daniel & John Broida

Available online: 20 Nov 2009

To cite this article: David B. Daniel & John Broida (2004): Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned, Teaching of Psychology, 31:3, 207-208

To link to this article: http://dx.doi.org/10.1207/s15328023top3103_6

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
All in all, requiring students to participate in research appears to confer modest benefits (or at least does not appear to cause harm). Most students derive some educational benefit from participating in research, report either positive or neutral reactions to their research experience, and would still choose to participate in research even when presented with an alternative that would require equal time and effort. However, if psychology departments are going to continue justifying required research participation in terms of educational benefit, perhaps researchers should strive to improve students’ research experiences. Ideally, students would report more than a modest educational benefit, and most students would report feeling positive rather than merely neutral about their research participation.

References


Note

Send correspondence to David Trafimow, New Mexico State University, Department of Psychology, MSC 3452, PO Box 30001, Las Cruces, NM 88003–8001; e-mail: trafimow@crl.nmsu.edu.

Encouraging Distributed Study: A Classroom Experiment on the Spacing Effect

William R. Balch
The Pennsylvania State University, Altoona

Two introductory psychology classes (N = 145) participated in a counterbalanced classroom experiment that demonstrated the spacing effect and, by analogy, the benefits of distributed study. After hearing words presented twice in either a massed or distributed manner, participants recalled the words and scored their recall protocols, reliably remembering more distributed than massed words. Posttest scores on a multiple-choice quiz covering points illustrated by the experiment averaged about twice the comparable pretest scores, indicating the effectiveness of the exercise in conveying content. Students’ subjective ratings suggested that the experiment helped convince them of the benefits of distributed study.

The spacing effect occurs when distributed study results in better memory than massed study does (e.g., Dempster, 1988). This effect holds true for both the spacing of items in a single list (e.g., Underwood, 1970) and the spacing of practice trials (e.g., Baddeley & Longman, 1978). Moreover, the effect applies to both short items like words (e.g., Verkoeijen, Rikers, & Schmidt, 2004) as well as more complex material like textbook passages (e.g., Kraft & Jenkins, 1981) and classroom lessons (Seabrook, Brown, & Solity, 2005).

By frequent testing, an instructor can effectively help students distribute their study and improve their grades (Fulkerson & Martin, 1981). A complementary method might be to use exercises illustrating the benefits of distributed study to motivate students to adopt this strategy by...
choice. I present a classroom experiment demonstrating the spacing effect.

Method

Participants

One hundred forty-five undergraduates (84 women and 61 men) at The Pennsylvania State University, Altoona, distributed among two introductory psychology classes, participated during their first class of the Fall 2005 semester.

Materials

I selected 16 experimental words from a random sample of 120 common nouns (Zechmeister & Nyberg, 1982, p. 350) originally from the Spreen and Schulz (1966) norms. The experimental words were message, basket, fashion, justice, artist, supper, ticket, remark (Subset A); and cousin, leather, witness, pattern, bottle, empire, giant, habit (Subset B). Every word contained two syllables and each word in Subset A was approximately matched—on the dimension of concreteness—with a word in Subset B.

Procedure

First, I instructed students that I would read a list of words for them to remember and that most of the words would occur twice somewhere on the list. After the last word, they would count backwards by threes from a two-digit number that I gave them, at a rate of 1 number per second. Then, when I said “Recall,” they would write down, in any order, as many words as they could remember.

Reading one word every 3 sec, I presented a list of 36 words to each class. In the first four positions were two-syllable buffer words—vessel, household, household, and tower, in that order—to control for the primacy effect. In the remaining 32 positions were two occurrences each of the 16 experimental words. I counterbalanced across the two classes the assignment of the aforementioned word subsets (A or B) to spacing conditions (massed or distributed).

To generate the order of experimental words, I assigned one number to each of the massed words (twice-occurring words occupying two consecutive positions) and two numbers to each of the distributed words (twice-occurring words with at least one other intervening word). Drawing the numbers randomly (with the constraints that neither the numbers of two different massed words nor the numbers of the same distributed word occurred consecutively) resulted in 0, or an average of 12, intervening positions between the occurrences of each massed or distributed word, respectively.

After presenting the words, pacing students through 18 sec of backward counting with a hand motion every second and giving them the signal to recall, I allowed 2 min for the recall task. Then, referring to a transparency of the massed and distributed words projected on a screen, students scored their recall protocols and indicated—by a show of hands—how many remembered more massed than distributed words, more distributed than massed words, or the same number of each word type.

In the debriefing, I identified the experimental hypothesis (people recall more distributed than massed words), the independent variable (massed or distributed words), the dependent variable (number of words recalled), and control procedures. These control procedures included the use of buffer words to eliminate the primacy effect, the backward counting task to eliminate the recency effect, and counterbalancing across classes to control for differences in the memorability of specific words.

To evaluate objectively the effectiveness of the experiment as a teaching tool, all students received a content quiz on the same class day as the experiment. This quiz—taken as either a pretest (Class 1) or posttest (Class 2)—consisted of 10 multiple-choice questions. Seven questions (on massed practice, the relation of massed practice to cramming, distributed practice, the use of buffer words, primacy, recency, and free recall) related specifically to the spacing-effect demonstration and three questions (on recognizing novel examples of an independent variable, a dependent variable, and a counterbalanced variable) related to experimental method in general. As a subjective evaluation, students rated the exercise on convincingness (how much it helped convince them of the benefits of distributing their study over time) and enjoyment.

Results

Spacing Effect

A one-way, within-participants ANOVA combined across classes revealed that students recalled significantly more distributed (47.8%) than massed words (34.5%), F(1, 144) = 44.74, p < .001, η² = .24 (see Table 1). This effect was significant for both women, t(83) = 5.98, p < .001, and men, t(60) = 3.29, p < .005.

Objective Evaluation

Students averaged 38.3% (Class 1) on the 10-question pretest and 77.9% (Class 2) on the identical posttest (see Table 1). A one-way, between-groups ANOVA revealed that this difference was significant, F(1, 143) = 308.04, p < .001, η² = .68. On the seven spacing-effect questions, students averaged 40.7% on the pretest versus 84.1% on the posttest, t(143) = 16.58, p < .001. The analogous scores on the three experimental method questions were 32.9% versus 67.1%, respectively, t(143) = 6.33, p < .001. Better posttest than pretest perfor-
mances suggest that the exercise effectively presented concepts related to research methodology in general as well as to material specifically related to the demonstration.

Subjective Evaluation

Averaged across classes, students rated the exercise 8.1 on convincingness and 7.2 on enjoyment, based on a scale from 0 (lowest) to 10 (highest; see Table 1).

Discussion

Although researchers are still trying to determine the best theory of the spacing effect (e.g., Toppino & Bloom, 2002), the robustness, reliability, and broad scope of this effect is clear (Dempster, 1988). Moreover, teachers have several likely means of improving their students’ academic performance through distributed study, including frequent testing (Fulkerson & Martin, 1981) and exercises illustrating the benefits of distributed study. Regarding the latter approach, I have extended the strategy of Wesely (1999), who assigned students the task of evaluating the effectiveness of distributed study (and other methods) in terms of evidence from the research literature. The rationale for my technique of demonstrating the spacing effect directly to students was to make the advantage of distributed study as real as possible to them.

By having students score their own recall protocols, I could illustrate the spacing effect immediately (albeit roughly) through a show of hands. About two thirds of the students personally showed the effect, remembering more distributed than massed words. On subsequent checking, I found that about 90% of the students had scored themselves correctly and that only 3% had made errors that changed their conclusions about whether they had obtained the effect. Comparing the average convincingness ratings of students who had shown the effect (8.14) with the ratings of those who had not (7.94), I found no significant difference, t(143) < 1, p > .10. Apparently, the students based their judgments on the overall results of their class.

My students’ relatively high convincingness ratings (M = 8.1 out of 10) directly after the demonstration, although encouraging, did not indicate what effect the demonstration may actually have had on students’ subsequent studying. To address this problem, I recently asked 79 additional introductory psychology students to rate both their pre- and postdemonstration study habits 11 weeks after participating in the spacing-effect demonstration during the Spring 2006 semester. They made a vertical mark on each of two horizontal-line scales labeled massed study at one end and distributed study at the other; then I converted their marks to a numerical scale from 0 (massed) to 10 (distributed). Their postdemonstration distributed-study rating (4.3) was significantly higher than their predemonstration rating (2.6), F(1, 78) = 137.07, p < .001, η² = .64. Because these ratings were subjective, retrospective, and possibly biased due to my asking for both ratings at the same time, they might not accurately represent students’ actual patterns of study. Yet to the extent that the ratings were informative, they suggest that the students tended to cram both before and after the demonstration but that the demonstration may have reduced their cramming slightly.

Instructors who try this demonstration should consider counterbalancing massed and distributed words across word subsets in two different classes and sharing the results with each class. In addition, instructors might discuss the relation between the spacing effect and the practice of cramming for tests. Finally, they might ask their students to suggest future experiments that address the following issues: testing the spacing effect on performance in a real college course (rather than on word recall), objectively assessing the influence of spacing effect demonstrations on students’ actual study habits, comparing the effectiveness of such demonstrations with that of a lecture on the same topic, and examining the possible effects of age and other participant variables on the distribution of study. In this way, instructors might stimulate a postdemonstration discussion integrating the spacing effect and general research methodology.

Table 1. Word Recall, Content Scores, and Subjective Ratings in Two Different Classes

<table>
<thead>
<tr>
<th>Sample</th>
<th>Word Recall</th>
<th>Content Score</th>
<th>Subjective Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distributed</td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td>Class 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>48.2</td>
<td>38.3</td>
<td>8.1</td>
</tr>
<tr>
<td>SD</td>
<td>18.2</td>
<td>17.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Class 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>47.5</td>
<td>77.9</td>
<td>8.0</td>
</tr>
<tr>
<td>SD</td>
<td>19.1</td>
<td>19.9</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Note. N = 145.

*Percentage of eight distributed versus eight massed words, varied within participants. bPercentage of 10 multiple-choice questions correctly answered on a content quiz. cBased on a scale from 0 (lowest) to 10 (highest). dEach class received either a pretest or a posttest. eη = 75, with Subset A words distributed and Subset B words massed. fη = 70, with Subset A words massed and Subset B words distributed.
References


Notes

1. I thank Cynthia Stewart and Althea Eaton for their assistance in conducting, and in scoring data from, the classroom experiment reported in this study.

2. Send correspondence to William R. Balch; Psychology Department; The Pennsylvania State University, Altoona; 300 Ivyside Park; Altoona, PA 16601; e-mail: wrb3@psu.edu.

The Elusive Definition of Outliers in Introductory Statistics Textbooks for Behavioral Sciences

Thomas P. Hogan and Kimberly Evalenko
University of Scranton

We examined treatment of outliers in 40 introductory statistics textbooks for behavioral science. The majority of texts (75%) included treatment of outliers. Other than agreement in defining outlier as an extreme data point, there was great diversity in treatment. Of the 30 texts including outliers, 11 presented both the univariate and bivariate cases, 15 treated only the univariate case, and 4 only the bivariate case. The texts included 7 different operational definitions of outliers; 16 texts provided no operational definition. A supplementary analysis of outliers in 3 statistical software packages showed more consistency, although not complete agreement. We offer recommendations for treatment of outliers in introductory statistics courses for behavioral science.

Following Tukey’s (1970, 1977) pioneering work on exploratory data analysis, the concept of an outlier has become well established in the statistical literature. Research articles often refer to the occurrence and disposition of outliers in data analysis. Most introductory statistics textbooks for the behavioral sciences, as shown in this study, include reference to outliers. In addition, widely used statistical software packages include procedures for identifying outliers.

Many topics in the introductory statistics course (e.g., standard deviations, Pearson correlations) have standardized definitions; textbooks differ mainly in pedagogical techniques (e.g., number and types of examples used) rather than in substance. Formulas for calculating most routine statistics are highly standardized in different textbooks. By contrast, beyond the common definition of an outlier as an aberrant or very unusual data point, we have noticed considerable variety in how textbooks treat this concept. We investigated how introductory textbooks on statistics in the behavioral sciences cover outliers. We supplemented the textbook analysis by analyzing outliers in three statistical software packages.

For purposes of reference, we constructed Figure 1 showing the principal terms used in Tukey’s (1970) original treatment of outliers. The figure applies Tukey’s terms to a normal distribution of T scores (μ = 50, σ = 10). To keep the figure size manageable, we included only even-numbered scores. Where important reference values in T scores or standard units occur at odd-numbered scores, we placed them at the nearest even-numbered score. The crucial parts of Tukey’s scheme are (a) use of the interquartile range (IQR) as the basic unit (rather than standard units) for measuring distances; (b) a multiplier of 1.5 applied to the IQR; and (c) a two-tiered system, going 1.5 × IQR beyond the first and third quartiles. Tukey’s original presentations did not include the term outlier. Rather, he labeled data points in the first tier as outside values and data points in the second tier as far out values. Today, use of the term outlier is nearly universal.

Method

We examined 40 introductory statistics textbooks (see Appendix) using several inclusion criteria. The text had to be aimed at the social and behavioral sciences; we did not include mathematical statistics books. The text had to be intro-
Improving Students’ Study Habits by Demonstrating the Mnemonic Benefits of Semantic Processing

Julie M. Bugg
Washington University

Edward L. DeLosh
Colorado State University

Mark A. McDaniel
Washington University

This article describes an in-class exercise that illustrates the advantage of semantic over nonsemantic study habits. The exercise includes a survey of students’ current study strategies, followed by the presentation of an abbreviated version of Craik and Tulving’s (1975) classic levels-of-processing experiment. We observed significant benefits of semantic processing over nonsemantic processing, and this result motivated an in-depth discussion regarding the limitations of students’ intuitions about effective study strategies and methods for improving current strategies. The brief exercise changed students’ intended strategies for future studying and helped students learn the concept of semantic processing.

Students often have misconceptions about the best strategies for committing information to memory. Informal surveys in our courses show that students try to memorize facts by repeating them over and over again and learn textbook information by reading the text multiple times. The memory literature shows that rote repetition is not particularly effective, however, especially relative to strategies that involve semantic or elaborative processing of information (Cain & McDaniel, 2007a; Craik & Watkins, 1973). The mnemonic benefit of semantic processing is, perhaps, one of the most widely held notions to come out of the memory literature. Nonetheless, students frequently persist in using a rote repetition strategy.

This exercise illustrates the robust mnemonic advantage of semantic over nonsemantic processing and encourages students to consider the implications of this advantage for their study habits. We based the exercise on an abbreviated form of Craik and Tulving’s (1975) classic levels-of-processing experiment demonstrating the benefit of semantic processing, presented in the context of an interactive discussion of students’ study habits. During the exercise, students reveal their beliefs that rote repetition is an effective study method, only to have that intuition challenged by subsequent demonstration and discussion.

Although similar exercises appear in textbooks (e.g., Francis, Neath, Mackewn, & Goldthwaithe, 2004; Neath, 1998) and in this journal (Chaffin & Hermann, 1983), our exercise is more extensive in providing a framework for the levels-of-processing experiment, surveying students about their study habits, providing data to support key arguments, offering discussion questions, and encouraging students to consider the implications for their studying.

Method

Participants and Procedure

Undergraduates in an experimental methods course (n = 51) participated in the exercise. We modeled the exercise after Craik and Tulving’s (1975) classic depth-of-processing experiment and presented it using
Orthographic processing, however, we note that the one-way ANOVA revealed that it fit in the sentence ‘The ___ was building a nest!’). We randomly intermixed the orienting questions and presented them for 3 sec each. Students answered each question by writing “yes” or “no” on an answer sheet numbered 1 to 18. We did not inform students that we would give them a memory test for the words.

After the instructor presented the words, she asked students to describe the different types of questions given and consider why they were included. This short discussion (2–3 min) served as an intervening activity. On the subsequent surprise free recall test, the instructor gave students 60 sec to recall the words on an answer sheet. Students subsequently scored their recall responses for each type of orienting question. The collection and scoring of data took approximately 7 min.

The instructor then asked students to reveal the pattern of results in their data by raising their hand to indicate which of the three conditions produced the highest and lowest levels of recall. Students readily determined the relative order of recall performance with the semantic processing condition leading to better recall than either the phonological or orthographic processing conditions. Next, the instructor led the class in the discussion and interpretation of the results. The instructor highlighted the findings of Craik and Watkins (1973) and Craik and Tulving (1975) showing that successful remembering is better achieved through the use of semantic study strategies rather than phonologically based or rote repetition strategies. Students then considered the implications of the exercise, discussing study strategies suggested by the research literature and the in-class exercise. Students proposed ideas such as relating information to existing knowledge, making material personally relevant, and thinking more deeply about a topic by generating examples and applications.

Results and Discussion

Before and after the exercise, students used a 5-point Likert scale ranging from 1 (not at all effective) to 5 (very effective) to rate the effectiveness of two study strategies: reading and rereading information versus relating information to oneself. Before the exercise, students reported the strategy they used to study for the previous exam in the course. After the exercise, students indicated the strategy they intended to use in preparing for the next exam. Options included repeating the material over and over again, reading and rereading the material, thinking about the meaning of the material, and applying the information to one’s personal experiences. We permitted students to describe a strategy that was not one of the aforementioned choices, and we classified the provided strategies as semantic or nonsemantic (e.g., rote repetition). If a student indicated use of both types, we classified the strategy as semantic. Finally, to assess students’ ability to apply semantic processing to real-world learning and memory situations, we administered the same four multiple-choice questions before and after the exercise (see Table 1).

<table>
<thead>
<tr>
<th>Table 1. Sample Multiple-Choice Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhonda just received a new banking card with a PIN number that she needs to remember to access her checking account. She is trying to devise a strategy for remembering the PIN number. Which of the following would you recommend?</td>
</tr>
<tr>
<td>a. Try to remember the way the numbers look, like whether they are more circular (like a 0) or more straight edged (like a 4)</td>
</tr>
<tr>
<td>b. Make up a rhyme that includes the numbers and try to recall the rhyme when you are at the ATM</td>
</tr>
<tr>
<td>c. Relate the numbers to something of personal relevance like the jersey numbers of your favorite sports stars</td>
</tr>
<tr>
<td>d. I would recommend any of the above because they are equally likely to help her remember the PIN number</td>
</tr>
</tbody>
</table>

We conducted a $2 \times 2$ repeated measures ANOVA for the effectiveness ratings with type of study strategy (reading and rereading vs. relating to oneself) and time (pre- vs. postexercise) as factors. Most critically, the two-way interaction was significant, $F(1, 50) = 42.31, p < .001$, partial $\eta^2 = .46$. Prior to the exercise, students rated reading and rereading information ($M = 3.88$) to be as effective as relating information to oneself ($M = 3.76$) as a study strategy. After the

\[ F(1, 100) = 88.44, p < .001. \]
exercise, students rated relating information to oneself (M = 4.31) as significantly more effective than reading and rereading information (M = 3.25), t(50) = −4.79, p < .001, d = 1.05. Before the exercise, only 14% of students reported using a semantic study strategy in preparing for their most recent exam. Following the exercise, 68% said they would use a semantic strategy for the next exam. Finally, students answered significantly more multiple-choice questions correctly after the exercise (M = 82%) than before the exercise (M = 49%), t(50) = −7.75, p < .001, d = 1.27.

These results converge on the conclusion that the exercise improved students' understanding of components yielding effective memory and study strategies. Moreover, the exercise was effective in changing students' reports of their intentions for future studying. If students' actual behaviors followed these intentions, then the exercise could enhance the effectiveness of students' study activities. Students report rereading as a preferred study method in this study (see also Karpicke, 2007), yet such a strategy appears to be minimally effective for improving retention and learning (Callender & McDaniel, 2007b). The vehicle for changing studying intentions included a discussion of the shortcomings of commonly used learning and memory strategies, as well as alternative strategies students might use to take advantage of depth-of-processing effects. This discussion was motivated by students' participation in an abbreviated levels-of-processing study replicating Craik and Tulving's (1975) finding of better memory for semantic than nonsemantic processing and completion of a class survey regarding study strategies.

One limitation is that we assessed students' learning and intentions immediately following the exercise, and we did not assess actual changes in strategy use. Nonetheless, the data do show that the brief and easily implemented exercise was effective for informing students about useful study strategies, convincing students of the need to improve their study habits, and for the initial learning of fundamental concepts from the memory literature.

References


Notes

1. The file for the abbreviated levels-of-processing experiment is available on the Internet at http://lamar.colostate.edu/~delosh/downloads.htm. This file was designed for use with Microsoft PowerPoint, versions 97 and later, on either Windows or Macintosh platforms. You can also present this experiment using Microsoft’s free PowerPoint Viewer, which is available at the same location or from Microsoft’s Web page (http://www.microsoft.com).

2. Send correspondence to Julie M. Bugg, Department of Psychology, Washington University, Campus Box 1125, 1 Brookings Dr., St. Louis, MO 63130; e-mail: jbugg@arts.wustl.edu.
References


Note

Send correspondence to Bruce Henderson, Department of Psychology, Western Carolina University, Cullowhee, NC 28723; e-mail: henderson@wcu.edu.

A One-Minute “Intelligence” Test

Richard A. Griggs

*University of Florida*

Testing, especially abilities testing, is an essential topic in psychology education because of its importance to students in both academics and the nonacademic world. In this article, I describe a class activity requiring minimal preparation that serves as an engaging stimulus for discussion of many important aspects of testing. Although I describe my use of the activity in introductory psychology, it has other uses in this course and in other psychology courses involving testing.

College students are clearly familiar with testing, both academic and standardized (e.g., ACT, SAT) tests. Similarly, they realize the importance of test performance to their success in academics and to their future careers. Given its importance, testing is a constant topic for public debate (e.g., the recent “Bell Curve” controversy; Herrnstein & Murray, 1994). Thus, it is essential for psychology teachers to provide students with a solid understanding of testing, especially abilities testing. A brief examination of courses and textbooks indicates that psychologists realize the importance of these topics for their students. Most schools have courses in tests and measurements; Messer, Griggs, and Jackson (1999) found that 82% of the schools surveyed offered such a course. In addition, intelligence and testing are topics covered in several other courses beginning with the introductory course. It is crucial to do a good job on these topics in the introductory course because this is the only experience most undergraduates have with psychology. I checked *A Compendium of Introductory Psychology Textbooks* (Jackson, Griggs, Marek, & Christopher, 1998) and found that all but one of the current full-length introductory texts have a chapter or major chapter section on intelligence and testing, providing further evidence of its importance in psychology. I designed the activity I describe in this article to open the unit in my introductory course on this topic.

In developing this activity 16 years ago, I reasoned that if students actually took a test that appeared on the surface to be biased, the experience would stimulate their thinking and facilitate learning about important aspects of testing such as different types of test bias and different types of validity (e.g., face validity vs. predictive validity). Because of the controversial nature of intelligence tests, a fictional intelligence test seemed ideal. However, I needed one that did not take much class time, could be scored easily, contained engaging test items, and could be administered in both small and large classes.

After a lecture on problem solving, an introductory student showed me a copy of *Omni Magazine: The Best Brainteasers From Omni Magazine* (Morris, 1983). It contained two “intelligence” tests (humorously labeled “College Entrance Exam 1” and “Advanced College Entrance Exam”) that were ideal for the class activity. These tests involved solving rebus-type puzzles, commonly called “wacky wordies” (see Figures 1 and 2). I used these two tests with only minor changes.

Procedure

The activity requires minimal preparation. The two tests (Figures 1 and 2) and the answer sheets (given in the Appendix) can be copied for individual administration and feedback or made into transparencies for group use.

There are two parts to the activity. Because students should not know that there will be two tests, the tests should not be numbered as they are here for identification purposes. Instruct students to number 1 through 24 on a sheet of paper on which

![Figure 1. Twenty-four test items in Test 1.](image-url)
they have to write their answers. Alternatively, if students receive individual copies of the test, they could write their answers on the test itself (I have found that this approach is not a good idea because most students want to reuse their test with other people). You should ensure that the students realize this is not an accepted intelligence test but only an informal exercise devised for a class activity. I also recommend that you mention this fact again after completing the activity.

The general pretest instructions (which you can read with the sample test item written on a blackboard) follow: “There are 24 items on the test. Each item is made up of letters, words, geometric shapes, and lines. Convert each to a verbal equivalent word or phrase. An example is: STTTHEORY.” Rather than giving the answer to this sample item, you should solicit potential answers and use the responses to further explain the nature of the test items. The answer to the sample item is “the inside story” (the word the inside the word story). This sample item is usually sufficient for student understanding of the nature of the test.

Allow 1 min for the first test. When time is up, tell the students to stop, count the number of answers they have written, and record that number at the top of their answer sheet. At this point, there are different ways to proceed, depending mainly on how much time you want to devote to the activity and whether you wish student performance to remain anonymous. For anonymity, merely collect the answer sheets without names on them. However, I have never had a problem with students worrying about anonymity in either large or small classes and thus have found the following path the most efficient with respect to time spent and knowledge gained. To get rough estimates of the central tendency and variance for the distribution of number of written responses for the class, I have everyone raise their hands. As I count aloud from 0 to 24, I have students put their hands down after I count one past their number of responses. It is fairly easy to visually determine the class central tendency and variability. Even with a 1-min time limit, the range is fairly large (usually 4 to 13 or so with a median around 6 or 7). You can use this informal analysis (or the results of a more formal one if you collect the test results) to review measures of central tendency and variance and problems with skewed distributions. Distributions for this test are usually skewed by a few high scores.

Next, go over the answers to the 24 items by having students provide the answers. Pause on some of the slightly more difficult ones (Items 11, 12, 19, 21, 22, and 23) and ask students who solved these items to provide clues for those who did not solve them. You may also relate these particular test items to the problem-solving section of the cognitive chapter (if you cover this material). The test items nicely illustrate fixation and invariably provide you with “live” classic “aha” cases of insight.

The first test is relatively easy and leads students to think that it is biased in that performance is more a function of time and motor skills (writing speed) than intelligence. Students are usually very involved and vocal about this impression. Guide the discussion to the question of how psychologists might decide whether this test, or any test, is valid. After a few minutes of such discussion, announce that the first test was just a warm-up exercise and that you will now give the real test. Have the students number 1 to 24 again and distribute (or show) Test 2. The students are very eager at this point, but this enthusiasm wanes quickly.

After a minute has elapsed, inform them and tell them that you will allow more time. At this point, the typical student has written only one or two responses and sees clearly that more is involved in solving these items than time and motor skill. The items on Test 2 are fairly difficult. I usually allow about 5 min for this test. Little is gained by more time. I check the central tendency and range again but provide only a few answers, which I solicit from students after they provide clues for the items. I inform the students that I will be providing a complete answer sheet later and do not want to deprive anyone of the joy of insight and solution.

Because both tests lead to substantial variability in performance, I usually briefly discuss the need for such variance in the calculation of correlations (another topic earlier in the course) and the use of correlations in calculating validity and reliability measures. Following this discussion, I explain test validity, stressing criterion-related predictive validity for abilities tests, and then relate predictive validity to the different senses of bias in testing. I point out that almost everyone agrees that intelligence tests are usually biased in requiring certain cultural experiences for optimal performance. I contrast this intuitive sense of bias, however, with bias in the statistical sense—different predictive validity for different ethnic groups (e.g., for a thorough discussion of this latter type of bias, see Anastasi & Urbina, 1997, chap. 6; for an introductory-level discussion of both types of bias, see Myers, 1998, pp. 359–360).1

1 As part of a package of simulations and demonstrations for introductory psychology, a former teaching assistant for my course, Sarah Randsell, developed an interactive computer module version of this activity (Randsell, 1992). Her module, however, focuses mainly on test reliability. Thus, the second test in the module is similar in level difficulty to Test 1 to check reliability, illustrating another direction a teacher could take with this activity. If you wish to develop additional tests in personalizing the activity for your own purposes, numerous additional “wacky wordy” puzzles are available in Doctor (1998).
To contrast these two types of possible biases, I pose the following hypothetical: Assume performance on this test is strongly positively correlated with performance on accepted intelligence tests and equally so for various ethnic groups. Given this assumption, what can you conclude about the validity and possible bias of the test? Discussion of this hypothetical usually leads students to understand that the term bias has multiple meanings and that a test can be biased in one sense and not others. This discussion also gives them a better understanding of predictive validity and its utility.

Discussion

With respect to student input on the value of this activity, I have used it in more than 50 sections of introductory psychology with enrollments ranging from 20 to 300 students. When I have collected evaluation data on the activity (mainly when I first began using it and periodically since), the students have always found it to be both interesting and informative and were unanimous that I should use it in future sections of the course.

In addition to getting students actively involved in thinking about testing and such important topics as bias in testing, this activity can be related to many other topics in the introductory course, ranging from statistics (e.g., measures of central tendency and variance and correlations) to problem solving (e.g., fixation and insight). With respect to this latter topic, a teacher who did not want to use the tests might want to use some of the items as problems to illustrate various aspects of problem solving. The activity also serves as a springboard for ensuing discussions of a wide array of topics concerning intelligence and testing. Use of this activity is also not limited to the introductory course. It can be used in many different courses that involve testing, especially abilities testing (possibly methods, developmental, cognitive, and tests and measurements courses). It takes little class time, and students have always found it to be both an interesting and valuable activity.

References


Appendix

Answers to Test 1

1. Sandbox
2. Man overboard
3. I understand
4. Reading between the lines
5. Long underwear
6. Crossroads
7. Downtown
8. Tricycle
9. Split-level
10. Three degrees below zero
11. Neon lights
12. Circles under the eyes
13. Highchair
14. Paradise
15. Touchdown
16. Six feet underground
17. Mind over matter
18. He’s beside himself
19. Backward glance
20. Life after death
21. GI overseas
22. Space program
23. See-through blouse
24. Just between you and me

Answers to Test 2

1. Split-second timing
2. A long letter from home
3. All between us is over now
4. Six of one, half a dozen of another
5. It’s a small world after all
6. Unfinished symphony
7. Blood is thicker than water
8. Seven-up
9. Condescending
10. Scrambled eggs
11. No two ways about it
12. Line up in alphabetical order
13. A gross injustice
14. The odds are overwhelming
15. He’s an exponent of capitalism
16. Astronaut
17. Ambiguous
18. A wolf in sheep’s clothing
19. Sailing, sailing, over the seven seas
20. Assassinate
21. For no apparent reason whatsoever
22. A little misunderstanding between friends
23. A bad spell of weather
24. He came out of nowhere

Teaching of Psychology

Copyright ©2000. All Rights Reserved.
Enhancing Student Motivation:  
Extensions From Job Enrichment Theory and Practice

Arvid J. Bloom  
Stefani L. Yorges  
Angela J. Ruhl  
West Chester University of Pennsylvania

Job enrichment—the practice of redesigning jobs to be more motivating—has received considerable attention in recent years. To explore classroom extensions of the underlying principles, we administered a slightly modified Job Diagnostic Survey (JDS; Hackman & Oldham, 1975) to 217 psychology students along with outcome scales assessing their course motivation, satisfaction, performance, absenteeism, interest, and desire to withdraw. The results showed significance for 4 of the 5 JDS predictors in accounting for the outcome measures. The degree to which students perceived course material to affect the lives of others was the most salient predictor. Based on the findings, we discuss the relative importance of the course enrichment dimensions.

Despite the attention devoted to redesigning boring jobs in the workplace to make them more motivating through "job enrichment" (Hackman & Oldham, 1976, 1980), surprisingly little theory and practice has been extended to college classrooms. To explore teaching applications, we sought to identify enrichment dimensions that predict student outcomes in psychology courses.

Hackman and Oldham's (1976, 1980) theory involves five core job dimensions that predict intrinsic work motivation, job satisfaction, and effectiveness. In their job characteristics model of work motivation, skill variety (V) refers to the use of different skills and talents. Task identity (I) involves producing identifiable pieces of work with visible outcomes. Task significance (S) reflects a perception by workers that doing the job benefits others. Autonomy (A) involves freedom from close supervision. Finally, feedback (F) reflects workers' ability to learn about their effectiveness directly from the work itself rather than from supervisors. In line with the requirements in Table 1, job enrichment involves raising the levels of job characteristics to enhance work outcomes.

Hackman and Oldham (1976, 1980) combined the core dimensions to indicate the overall motivating potential of a job according to the formula \((V + I + S)/3 \times A \times F\), with higher scores associated with more intrinsically motivating jobs. Furthermore, they posited that workers' growth need

---

### Table 1. Components of Job Motivation and Proposed Analogues for Course Motivation

<table>
<thead>
<tr>
<th>Core Enrichment Dimension</th>
<th>Existing Job Characteristics Model</th>
<th>Proposed Course Characteristics Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill Variety</td>
<td>Different skills and talents are required by the job.</td>
<td>A variety of student knowledge and skills is required to learn the course material.</td>
</tr>
<tr>
<td>Task Identity</td>
<td>Workers produce identifiable pieces of work with visible outcomes.</td>
<td>Students make unique and identifiable classroom contributions.</td>
</tr>
<tr>
<td>Task Significance</td>
<td>Workers perceive that doing the job benefits the lives of others.</td>
<td>Students perceive that learning the material is likely to eventually benefit the lives of others.</td>
</tr>
<tr>
<td>Autonomy</td>
<td>Workers have the freedom to schedule their own tasks and to determine how to carry them out.</td>
<td>Students have the freedom to decide how to learn the material and fulfill the course requirements within broad parameters.</td>
</tr>
<tr>
<td>Feedback</td>
<td>Workers receive task-generated information about their job effectiveness, without reliance upon a supervisor or manager.</td>
<td>Students are able to assess their own grasp of the course material, without reliance on exams, grades, or other feedback from the instructor.</td>
</tr>
</tbody>
</table>

---

Participants consisted of 217 students (29% men, 71% women; M age = 21.9 years; 60% psychology majors) from nine psychology courses taught by seven different instructors at a medium-sized state university. The courses reflected a wide variety of instructional methods.

Participants anonymously completed 15-min written surveys during regularly scheduled classes. The surveys began with our 15-item classroom adaptation of Hackman and Oldham's (1975) Job Diagnostic Survey (JDS), a widely employed organizational research tool that assesses each core job characteristic. The classroom version preserved as much of the meaning of the original dimensions and scale anchors as possible. Two- to 4-item Likert-type scales ranging from 1 (strongly disagree) to 7 (strongly agree) assessed six student outcome measures: motivation ("I feel motivated to excel in..."),...
served across different evaluation questions. Future research on associations between expected grade and evaluations of teaching should examine in more depth what question content areas are most likely to be affected and explore possible explanations for differential question effects. One promising approach is to examine student evaluations within the context of the existing knowledge of impression management theory and exit interviewing.

References


Note

Send correspondence to Elizabeth M. Ginexi, who is now at Westat, 1650 Research Boulevard, Rockville, MD 20850; e-mail: ElizabethGinexi@Westat.com.

A Psychic-Reading Demonstration Designed to Encourage Critical Thinking

Timothy J. Lawson
College of Mount St. Joseph

Self-proclaimed psychics who perform seemingly accurate readings have been the focus of a number of recent television shows. I describe a psychic-reading demonstration—in which a professor performs a surprisingly accurate reading with a student in class—designed to encourage students to think critically about such readings. Students reported that they found the demonstration to be interesting, informative, and helpful for thinking about alternative explanations for the apparent accuracy of psychic readings.

People's strong interest and long-standing belief in questionable pseudoscientific and paranormal phenomena have prompted a number of psychologists to develop strategies for using such phenomena to teach students critical thinking skills (e.g., Banziger, 1983; Bates, 1991; Lilienfeld, Lohr, & Morier, 2001; McBurney, 1976; Wesp & Montgomery, 1998). These authors described ideas ranging from classroom demonstrations of telepathy to entire courses on pseudoscientific and paranormal topics.

Recently, a number of self-proclaimed psychic readers (e.g., Sylvia Browne, John Edward, James Van Praagh), some of whom supposedly communicate with the dead, have been the subject of a number of television shows and magazine articles (e.g., Christopher, 2001; Jaroff, 2001; Wiseman & O'Keefe, 2001). Seeing a reader reveal accurate and sometimes detailed personal information about a target person who is apparently a stranger may lead observers to infer that the reader is psychic. To encourage students to think critically about such readings, I developed a psychic-reading demonstration, including follow-up discussion questions. Although previous authors (e.g., Bates, 1991; McBurney, 1976) have described ideas for demonstrating telepathic abilities in the classroom (e.g., by guessing the playing cards or numbers selected by students), they did not describe how to perform a psychic reading that involves revealing specific details about a person one does not know well.

My demonstration relies on “cold-reading” (e.g., making general statements that apply to most people) and “hot-reading” (e.g., obtaining specific details about the person in advance) techniques similar to those apparently used by famous readers (see Jaroff, 2001; Nickell, 2001; Randi,


Textbook Coverage of Ethical Considerations in Research With Children

Kim Ernst
Loyola University New Orleans

I examined research methods, child development, and developmental psychology textbooks (N = 74) for ethical considerations concerning child research participants. Coverage varied both within and between textbook types, with many texts showing little or no discussion, especially of debriefing and recruiting incentives. Greater textbook attention to child participants would benefit students by raising their awareness of ethical considerations designed to safeguard children’s rights and welfare.

Educating psychology majors about research ethics is an essential component of the undergraduate curriculum (Brewer et al., 1993). Accordingly, course instructors have developed teaching methodologies that introduce students to ethical principles and teaching techniques to integrate this information into psychology course content (e.g., Beins, 1993; Matthews, 1991).

Teachers of psychology also have investigated textbook coverage of research ethics because textbooks are a significant source of information for students (McKeachie & Hofer, 2001). Two early studies revealed that coverage of research ethics was missing in 44% (n = 18) of experimental (Adair, Lindsay, & Carlropio, 1983) and 21% (n = 29) of introductory psychology textbooks (Korn, 1984). More recently, however, Fisher and Kuther (1997) reported that each of 14 introductory psychology textbooks mentioned research ethics. Nonetheless, results from introductory textbooks may not generalize to other psychology textbooks, and none of the studies previously mentioned explicitly addressed research ethics concerning child participants.

Federal laws and ethical codes promulgated by professional organizations mandate that investigators protect all research participants from physical and psychological harm. Researchers, therefore, should apply these standards of conduct to evaluate the ethical acceptability of their research. Ethical codes, however, are not exhaustive standards of conduct; instead, some ethical codes complement each other. Thompson (1990) and Miller (1998) asserted that the Ethical Standards for Research With Children (Society for Research in Child Development [SRCD], 1993) provide a more sensitive appraisal of children’s vulnerabilities than the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association [APA], 1992). Thus, it is common practice to apply more than one set of ethical guidelines when planning and conducting research (Miller, 1998).

Haemmerlie and Matthews (1988) suggested that coverage of ethical principles should occur in at least one required psychology course, although undergraduate programs have not standardized which course satisfies this recommendation. The research methods (RM) course seems best suited to fulfill this requirement because it is a required course of most psychology majors (Messer, Griggs, & Jackson, 1999), and it is more directly concerned with research ethics than other courses in the undergraduate curriculum. Moreover, social scientists consider research ethics a necessary topic of the RM course (Giesbrecht, Sell, Scialfa, Sandals, & Ehlers, 1997). As such, expecting RM textbooks to include ethical considerations for child participants seems plausible. One also might expect to find coverage of child participants in undergraduate textbooks for child development (CD) and developmental psychology (DV) courses. To date, however, textbook coverage of child participants has received no attention. Thus the purpose of this investigation was to determine the extent to which RM, CD, and DV textbooks present ethical considerations in research with children.

Method

Sample of Textbooks

I obtained textbooks copyrighted from 1995 through 1999 from colleagues and publishers. The time frame matched other current textbook research (e.g., Jackson, Lugo, & Griggs, 2001), and it allowed the inclusion of the most recent textbooks available at the time of this study. Excluded were textbooks published outside the United States, combined RM and statistics textbooks, student laboratory manuals, handbooks, study guides, and instructor’s manuals. The sample (N = 74) resulted in 30 RM, 27 CD, and 17 DV textbooks (references for the 74 texts are available on request).
Teaching the Principles of Test Validation in Introductory Psychology

Richard Wesp and Sussie Eshun
East Stroudsburg University of Pennsylvania

We describe a simple classroom activity that demonstrated the process of assessing predictive validity. We assessed changes in student understanding of the process, confidence in psychological testing, and enjoyment in participating in the exercise. Students who completed the activity performed better on a test question about validity and reported more confidence in testing. Students reported they enjoyed the exercise and that the activity provided them with a better understanding of the test development process.

Students have told us that they believe that psychological test questions were not good predictors of the attributes they were intended to predict. Their skepticism may originate in their belief that tests are simply a collection of questions that intuitively relate to what they are testing (Camac & Camac, 1993). Student misunderstanding of test construction and validation procedures have likely contributed to their lack of confidence in tests. Given the pervasiveness of psychological testing, students could benefit from an understanding of test development procedures. Textbooks discuss test construction; however, students can better understand it with practical experience.

Several papers have described activities that teach the science of psychological testing. Camac and Camac (1993) described a 6-week laboratory project in which experimental psychology students developed and evaluated the validity of a test for sensation seeking. Students completing the course demonstrated test development skills, rated the project as their most preferred laboratory activity, and agreed the activity effectively taught research methods.

In a similar project described by Hynan and Foster (1997), undergraduates in a psychological measurement class assessed reliability and construct validity of a test students designed and administered in class. Students made positive comments about the exercise and suggested that it helped them understand test development.

Reinehr (1991) described a simpler procedure for introductory-level students. Students assessed the instructor’s personality using a standard personality inventory and then compared their assessment with the instructor’s self-assessment. Reinehr used the results as the focus of class discussion of test construction, validity, and test use. He reported improved participation in discussion, a better grasp of testing in written work, and higher test grades.

Faculty might resist Reinehr’s (1991) activity because it is relatively complex and it requires self-disclosure. In this study, we introduced a simpler classroom activity on test development designed for nonmajors with little background in statistics. We assessed student enjoyment of the exercise, learning about the process of assessing predictive validity, and changes in confidence regarding testing.

Method

Participants

We used two comparable day classes of introductory psychology to assess the influence of a test-construction activity on student understanding of test development procedures. We assessed opinions about the activity in an evening class during a subsequent semester. Students were not psychology majors. The first author taught all classes with the same syllabus.

We randomly assigned one day class (n = 40) to participate in a test validity activity, and the other day class (n = 41) served as a discussion-only comparison class. We compared final grades of the test validity (n = 39, M = 79.5%, SD = 14.3) and the discussion-only class (n = 41, M = 80.8%, SD = 9.9), and found no significant difference, t(78) = 0.47, p = .64, Cohen’s d = 0.11; we concluded that the latter class was an appropriate comparison group.

Test Validity Activity Procedure

In the test validity activity class, the instructor introduced the concept of testing as part of a section on personality, explaining that researchers often design tests to make predictions about individuals. The instructor explained that for most tests, a researcher developed the test to measure some outcome. He then informed students about the day’s class activity, which involved identifying a characteristic about class members and then developing test items that might discriminate between those with and those without that characteristic.

Identify a characteristic to study. First, the instructor asked students to help identify characteristics of class members that differed dichotomously. The instructor suggested several examples such as athlete or nonathlete and players or nonplayers of musical instruments. Next, he tried to identify one dimension for which the class was relatively balanced (approximately 50/50 split) by asking class members to raise hands if they had certain characteristics (e.g., “play an instrument”). We found a relatively equal split between first- and later-borns.
Groups construct a question. Next, the instructor directed students to move into the 10 groups of 3 to 5 in which they had previously worked. He told groups that they had 5 min to select a question that had a dichotomous answer that might serve to discriminate between first- and later-borns. The instructor described several questions appropriate for the musical instrument example to help students understand what types of questions were acceptable. The instructor told students that questions could pose a statement that required an opinion as an answer (e.g., I like music: agree or disagree) or a forced choice (e.g., I prefer movies or concerts: movies or concerts). He told students that they must select a question that assessed a personality characteristic and was not physically tied to the answer (e.g., I have no older siblings). The instructor reminded students that their question should elicit different responses from first- and later-borns.

The instructor told students that when the group had an idea they should get instructor approval and then write their question and two alternative answers, a and b, on the blackboard. The instructor assisted groups, and when groups described to him an acceptable question, he announced it to the class to avoid duplicate questions. One group asked, “Did your parents give you a lot of responsibilities? (a) yes (b) no.”

Take the test and record answers. When all of the group questions were on the blackboard, the instructor numbered them and told all class members to write their answers to all questions. Next, he asked all of the first-borns only to identify by raising hands how they answered each question. He recorded responses, and then he repeated the procedure for the other questions. Next, he repeated both questioning and recording the same way with later-born students.

Assess findings. Last, the instructor asked students to identify patterns that indicated which questions were good predictors of birth order. Students first focused on questions in which both first- and later-borns responded strongly in the same way, failing to recognize that the solution required a comparison of the relative responses of the two groups.

The instructor prompted students to eliminate questions that did not serve to discriminate between conditions. About a third of the questions generated nearly identical responses from both groups, and another third of the questions were only moderately good predictors. We discarded them with an explanation that they were not useful because they did not help identify who had the characteristic. For each of the remaining questions, we calculated the percentage of students in each condition who answered “a” and compared the difference in percentage between conditions. We identified the question with the largest difference as the best and recognized that others could be useful questions. For example, if 9 of 18 first-borns agreed that their parents gave them a lot of responsibilities and 4 of 20 nonfirst-borns agreed, we converted to percentage scores of 50% and 20%, respectively. We then compared the difference of 30% with differences of scores for other questions to find the question with the largest between-group difference.

In a summary discussion of the activity, the instructor emphasized that the validation procedure identified those questions that highlighted the difference between first- and later-borns. He offered examples of how to use the same procedure with different types of tests. The instructor asked students to speculate why some questions failed to show a difference. The instructor prompted students to recognize ambiguity in how questions were worded or constructs defined. He discussed how none of the questions perfectly discriminated between first- and later-borns. Then he explained how multiple measures could compensate for having individual test questions that were only moderately good predictors.

Comparison Class Procedure

The instructor reviewed the same issues in the comparison class. Rather than conducting the activity, he used concrete examples of how to assess validity by comparing answers of different criterion groups on specific test questions.

Assessment

Test performance. We assessed the influence of the activity on student understanding of the process of test validation using a multiple-choice item on an exam. The exam question asked, “To create a test for a psychological characteristic, one should ____.” One answer was a nonsense foil, a second described the view that questions are simply based on the test developer’s beliefs, a third answer suggested the idea of item discrimiability but without reference to empirical assessment, and a fourth answer described an empirical approach to validating test items. We include the verbatim answers in Table 1.

Confidence in testing. Students rated their confidence in the results of tests designed to assess psychological disorders in a pretest on the third day of class and a posttest at the final exam. Students rated confidence on a scale from 1 (not at all confident) to 7 (very confident). We assessed the influence of the activity on student confidence in tests by comparing changes in confidence ratings between classes. Forty-one students in the comparison and 39 in the activity class completed both surveys.

Table 1. Percentage of Students in Each Class Selecting Alternative Exam Concept About Test Item Selection

<table>
<thead>
<tr>
<th>Conceptual Answers</th>
<th>Validity Activity</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical approach to validating the items</td>
<td>63.4</td>
<td>47.8</td>
</tr>
<tr>
<td>Item discrimibility sans empirical assessment</td>
<td>29.3</td>
<td>22.3</td>
</tr>
<tr>
<td>Nonsense foil</td>
<td>7.3</td>
<td>10.0</td>
</tr>
<tr>
<td>Intuition of the test developer</td>
<td>0.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

*Answer was: Give a test to people who are known to have the characteristic and those who do not and identify questions that discriminate between those who have the characteristic and those that do not. *Answer was: Develop a test that includes many questions that would seem to discriminate between those who have the characteristic and those that do not. *Answer was: Give a test to people and identify those questions that answers vary the most. *Answer was: Write many questions that seem to detect the characteristic.
Opinion ratings. In a subsequent semester, we conducted the demonstration in an evening class and asked students to rate the activity. Eighteen students rated agreement with the statements “I believe that I have a much better understanding of how tests are constructed,” and “I enjoyed the exercise.” Students rated each statement on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree).

Results

Learning

We compared the distribution of answers on the multiple-choice question on testing in the two classes (see Table 1). A Fisher’s exact test showed the distribution of answers differed significantly from chance ($p = .014$). An assessment of cell values showed that discrepancies in the portion of students selecting the correct answer (63% and 48% in the activity and comparison class, respectively) and portion believing that the test maker intuitively selects questions (0% and 20% in the activity and comparison class, respectively) caused most of the differences. A significantly larger portion of those in the activity class (93%) than those in the comparison class (70%) selected one of the two answers that referred to discriminability ($p = .011$, Fisher’s exact test).

Student opinions about the influence of the exercise on their understanding of test construction supported examination results. Students agreed ($M = 4.17, SD = 0.51$) that the activity improved their understanding.

Confidence in Testing

Students who participated in the activity had a small but significantly higher increase in confidence in tests than the comparison class. Our comparison of change scores showed student confidence rose in the activity class ($M = 0.2, SD = 1.4$) and fell in the comparison class ($M = –0.3, SD = 1.0$); the difference in change scores was significant, $t(78) = 2.03, p < .05$, Cohen’s $d = 0.41$.

Enjoyment

Students agreed that they enjoyed the activity. They gave the exercise a mean rating of $4.33 (SD = 0.49)$.

Discussion

We employed a very basic comparison procedure to actively demonstrate test-item validity without using traditional statistical assessment. Students enjoyed the activity, developed a better understanding of test construction, and became more confident in testing. We believe that our mathematically friendly procedure reflected the intent of traditional tests of predictive validity and better communicated the concept of validity to students with little background in statistics. Instructors might supplement our procedure with traditional statistics and should consider more sophisticated activities (e.g., Camac & Camac, 1993) for advanced courses.

We believe that our testing and opinion data provided a convincing demonstration of the activity’s overall effectiveness. Subsequent research might use additional exam questions to separately analyze the influence of the activity on student understanding of test construction and assessment. Also, future studies should eliminate from analysis students who were absent on the activity day. We did not keep attendance, but absences rarely exceeded three students; we believe that their elimination would have had little influence on our findings.

We have described in this article a relatively short (under 30 min), enjoyable, and simple test-construction activity. After using it many times, we are confident that it effectively improves student understanding of test construction and validation.

References


Notes

1. We thank Adrienne Lee and three anonymous reviewers for their valuable comments on an earlier draft of this article.

2. Send correspondence to Richard Wesp, Department of Psychology, East Stroudsburg University, 200 Prospect Street, East Stroudsburg, PA 18301; e-mail: rkwesp@po-box.esu.edu.

---

Reviewer 2 asked our opinion about the influence of the length of answers on student selection of answers (the correct answer was the longest). Both classes had the same questions so the influence was balanced across conditions. The influence of the activity may have combined with, but overpowered, the possible influence of question length. Twenty percent of those in the comparison class selected the shortest answer, and none in the activity class did.
Introducing Psychology Students to Research Methodology: A Word-Pleasantness Experiment

William R. Balch
The Pennsylvania State University, Altoona

This classroom experiment demonstrates research methodology and introduces cognitive (i.e., memory) as well as affective (i.e., pleasantness) topics. As an incidental task, introductory psychology students rated a series of pleasant and unpleasant words that were comparable on several other semantic dimensions. Based on data from 4 different classes, students recalled significantly more pleasant than unpleasant words. Students also scored higher on a posttest of content illustrated by the experiment compared to students receiving a pretest of the same content. More generally, the experiment provides a springboard for discussing the relation between experimental control and research ethics.

Students rate research methods as one of the least interesting topics in the introductory psychology course (Zanich & Grover, 1989). To address their lack of enthusiasm about methodology, I performed and evaluated a classroom experiment illustrating research methods and drawing on two topics that students rate as more interesting: memory and emotion. This experiment tests an effect predicted by the Pollyanna principle (Matlin, 2004; Matlin & Stang, 1978) and demonstrates that people recall more pleasant than unpleasant items. Although sometimes controversial (e.g., Mayer, 1986), this effect usually holds true for both words (Matlin, 2005, p. 140) and autobiographical events (Walker, Skowronski, & Thompson, 2003). The rationale for my experiment was to provide a teaching tool for introducing experimental methods as well as examples of both cognitive (i.e., memory) and affective (i.e., word-pleasantness) topics.

Method

Participants

Two hundred eighty-five undergraduates (163 women, 122 men) at The Pennsylvania State University, Altoona, distributed among four introductory psychology classes, participated. The experiment occurred during the first week of the Fall 2004 (Classes 1 and 2) or Spring 2005 semesters (Classes 3 and 4), before students received any information (in class lectures or through reading assignments) about research methodology or other content areas included in the experiment.

Materials

I selected the words from the word norms of Toglia and Battig (1978), with ratings based on scales from 1 (lowest) to 7 (highest). The 12 pleasant words were hope, style, interest, cure, beauty, trust, liberty, comfort, benefit, praise, travel, and wisdom (M = 5.5, SD = .4 on the pleasantness scale). The 12 unpleasant words were fool, quarrel, hunger, loss, theft, decay, trouble, insult, panic, grudge, fraud, andumble (M = 2.6, SD = .3). The pleasant and unpleasant words were about equal in intensity. The pleasant words averaged 1.5 points above the midpoint (4.0) of the pleasantness scale, and the unpleasant words averaged 1.4 points below the midpoint. To reduce the primacy effect, two neutral buffer words appeared at the beginning of the word list—area and meeting (with mean pleasantness ratings of 4.0 and 4.2, respectively).

All the words were comparable on four other scales (concreteness M = 3.5; imagery, M = 4.3; meaningfulness, M = 4.8; familiarity, M = 6.1). However, two or three of the pleasant words (depending on the pronunciation of the word interest) contained three syllables, whereas none of the unpleasant words did. Yet this slightly longer average length of the pleasant words would be unlikely to affect recall because the retention interval in this experiment exceeded the span of working memory. Moreover, any possible influence of word length would presumably favor the unpleasant words and work against the effect predicted by the Pollyanna principle.

Procedure

Hearing one word every 6 sec, students rated each word by circling a number on a pleasantness scale from 1 (lowest) to 7 (highest), in rating booklets with 10 pleasantness scales appearing on each page. The two buffer words, followed by two consecutive random orders of the 24 experimental words, made a total of 50 word ratings.

Then, to prevent a recency effect, participants took 3 min to complete an arbitrary filler task questionnaire consisting of 12 questions on their academic background.

For an unexpected recall task, they wrote—on the blank back of the last page of the questionnaire—all the words they could remember from the rating task, in any order. (They did not need to recall the ratings that they had given the words.) After 2 min, I collected their booklets and debriefed them on the experiment.

In the debriefing, I identified the experimental hypothesis (people recall more pleasant than unpleasant words), the independent variable (high- vs. low-pleasantness words), the dependent variable (number of words recalled), and controlled variables (concreteness, imagery, meaningfulness, and familiarity). Although I pointed out that unexpected word characteristics could possibly have influenced the re-
Results, the characteristics controlled in this experiment are the most important for word-memory experiments (e.g., Mayer, 1986). In addition, I mentioned that the buffer words at the beginning of the word list and the filler task at the end controlled for primacy and recency effects, respectively.

To assess the effectiveness of the experiment as a teaching tool, all four classes received a content quiz consisting of 10 multiple-choice questions that occurred on the same class day as the experiment. This quiz—taken as either a pretest (Classes 1 and 3) or posttest (Classes 2 and 4)—required students to recognize an example of an independent variable, a dependent variable, and a controlled variable, as well as to answer questions on research ethics (informed consent and debriefing), on the meaning of cognition and affect, on serial position effects (primacy and recency), and on the recall method of testing memory. Finally, students subjectively rated the experiment on enjoyment and on the perceived amount of their learning from the experiment, using scales from 0 (lowest) to 10 (highest).

Results

Word Pleasantness

Table 1 shows the recall performance of each class. Combined across classes, students recalled more pleasant (35.9%) than unpleasant words (30.9%). A one-way, within-participants ANOVA revealed that the word-pleasantness effect was significant, \( F(1, 284) = 28.83, p < .001, \eta^2 = .09. \)

The word-pleasantness ratings were highly reliable, as indicated by the correlation between students’ mean ratings of the 26 words and the ratings listed in the Toglia and Battig (1978) norms, \( r(24) = .95, p < .001. \) In addition, the average ratings of the pleasant (5.5) and unpleasant (2.5) words corresponded closely to the analogous ratings in those norms (5.5 and 2.6, respectively).

Content Quiz

Students averaged 56.1% (across Classes 1 and 3) on the pretest and 83.2% (across Classes 2 and 4) on the posttest (see Table 1). This difference was significant, \( t(283) = 15.26, p < .001 \) (between-groups \( t \)). Better posttest than pretest performance suggests that the exercise effectively presented concepts such as experimental method, research ethics, memory as a cognitive topic, and word pleasantness as an affective topic.

Subjective Ratings

Averaged across all classes, students rated the exercise 7.9 on perceived learning and 7.3 on enjoyment (see Table 1), based on a scale ranging from 0 (lowest) to 10 (highest).

Discussion

Although I developed this classroom experiment for use at the beginning of introductory psychology courses, it might also be appropriate for courses such as cognitive psychology, memory, or research methods. In presenting the words, instructors might either read them to the students, as I did, or show the words visually by projecting them on a screen at the same rate of 6 sec per word. Not including the content-quiz and subjective-rating evaluations, the entire exercise takes about 25 min. By reading back the pleasant versus unpleasant

<table>
<thead>
<tr>
<th>Sample</th>
<th>Word Recall(^a)</th>
<th>Content Score(^b)</th>
<th>Subjective Rating(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pleasant</td>
<td>Unpleasant</td>
<td>Pretest(^d)</td>
</tr>
<tr>
<td>Class 1(^a)</td>
<td>M</td>
<td>32.1</td>
<td>27.3</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>13.4</td>
<td>15.1</td>
</tr>
<tr>
<td>Class 2(^b)</td>
<td>M</td>
<td>38.0</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>13.6</td>
<td>12.2</td>
</tr>
<tr>
<td>Class 3(^c)</td>
<td>M</td>
<td>35.6</td>
<td>31.5</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>16.7</td>
<td>15.7</td>
</tr>
<tr>
<td>Class 4(^d)</td>
<td>M</td>
<td>37.7</td>
<td>32.5</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>12.7</td>
<td>13.7</td>
</tr>
</tbody>
</table>

Note. For all classes combined, \( N = 285. \)

\(^a\)Based on percentage of 12 pleasant versus 12 unpleasant words, varied within participants. \(^b\)Based on percentage of 10 multiple-choice questions correctly answered on a content quiz. \(^c\)Based on a scale ranging from 0 (lowest) to 10 (highest). \(^d\)Each class received either a pretest or a posttest. \(^e\)\( n = 68. \) \(^f\)\( n = 69. \) \(^g\)\( n = 68. \) \(^h\)\( n = 80. \)
words afterward and letting students score themselves, instructors could quickly indicate on the blackboard—from a show of hands—how many students recalled more pleasant than unpleasant words, more unpleasant than pleasant words, or the same number of each word type.

Along with the debriefing described in the Procedure section, this exercise introduced the experimental method and illustrated cognitive topics like memory as well as affective topics like pleasantness. In addition, the experiment provided a relatively simple example of the complex relation between emotion and cognition. Another example of this relation that might interest students is cognitive therapy, based on the hypothesis that different emotions are related to different patterns of thought (Beck, 1976; Deffenbacher, 1990).

To address research ethics, I had students sign an illustrative informed consent form, although this procedure is not generally required for classroom exercises. In the debriefing, I mentioned that researchers sometimes withhold information from participants. For instance, I did not tell students about the memory test in advance. I used this point to stimulate a discussion of both the advantages and the ethical issues involved in experimental manipulation. Thus, classroom experiments such as this one provide the opportunity for instructors to discuss broader ethical principles as well as more specific topics.

References


Note

Send correspondence to William R. Balch, Psychology Department, The Pennsylvania State University, Altoona, 3000 Ivyside Park, Altoona, PA 16601; e-mail: wrb3@psu.edu.
Internal and external validity are key concepts in understanding the scientific method and fostering critical thinking. This article describes a class demonstration of a “botched” experiment to teach validity to undergraduates. Psychology students (N = 75) completed assessments at the beginning of the semester, prior to and immediately following the demonstration, and on an examination. Results indicated that students significantly increased quiz scores across 3 assessment points, scored higher on experimental validity questions on an exam, and increased use of critical thinking skills in a novel application. The demonstration engaged students during class and increased comprehension on the topic of experimental validity.

Teaching scientific methodology unifies themes across diverse content areas within psychology (Brewer et al., 1993). Students of psychology must obtain the knowledge base to flexibly apply scientific principles to promote improved analytical (Wolfe, Reynolds, & Krantz, 2002) and critical thinking skills (Halonen et al., 2003; Halpern, 1998). Concepts of internal and external validity (Shadish, Cook, & Campbell, 2002) are especially important to understanding empirical methodology. In addition, both concepts involve students as active collaborators in their learning because these concepts ultimately involve identifying and articulating contextual influences. In fact, in-depth understanding of internal validity (i.e., assessing the degree to which experimenters’ methodological operations match their conclusions) and external validity (i.e., assessing the match between sampling and generalizations drawn from this sampling) necessarily involve the five interrelated skill sets (e.g., argument analysis, hypothesis testing) identified by Halpern (1998) as essential to critical thinking.

Despite the enthusiasm of educators, students indicate modest interest in research and a preference for relatively passive class activities (Vittengl et al., 2004). A number of papers have described activities that teach various research topics aimed to reduce student passivity. For example, students participated in exercises to evaluate research hypotheses (Johnson, 1996; Sagarin & Lawler-Sagarin, 2005; Ward & Grasha, 1986), operational definitions (Herringer, 2000), and pseudo-scientific claims (Boyce & Geller, 2002). Several papers described experiments conducted within class to demonstrate research methodology, such as examination of pulse rate variations (Carr & Austin, 1997), haptic perceptions (O’Dell & Hoyert, 2002), and smiling frequency (Lipsitz, 2000). Overall these procedures engaged students in multiple aspects of the scientific method. However, all but one of these demonstrations involved students in advanced psychology classes, and no techniques focused specifically on topics of internal and external validity.

This article addresses these issues by describing a demonstration aimed at engaging introductory-level students in evaluating aspects of experimental validity. I evaluated the efficacy of a research demonstration intentionally riddled with serious threats to internal and external validity. Halpern and Desrochers (2005) suggested that “botched” classroom demonstrations can enhance active participation of students, a technique not yet used for teaching validity. In this
demonstration, students acted as experimental participants and observers and identified experimental confounds in a “botched” research demonstration. The presentation countered passivity in that students served as participants and critical judges. It aimed to hone students’ critical thinking skills, increase their comprehension of experimental validity, and stimulate their process of inquiry in psychological science.

Method

Participants

Undergraduates attending a general psychology course at a regional university campus over three consecutive semesters served as participants. An average of 80% of enrolled students attended the class demonstration, resulting in 75 participants.

Procedure

Students completed assessments at four time points to evaluate learning that resulted from the demonstration. All questions specifically measured student knowledge of internal and external validity in a multiple-choice format. Participants completed five questions within a larger mock quiz used to stimulate discussion about course content at the beginning of the semester. Immediately prior to the demonstration, students completed three quiz questions. At the end of the class in which the demonstration occurred, students completed six questions as part of a review exercise to consolidate learning. Four weeks following the demonstration, students completed an examination in which nine questions pertained to experimental validity. The remaining 51 questions related to developmental psychology. Students completed all questions as part of the overall requirements for the course. I selected questions from the test bank (Cooper, 2005) provided by the publisher of the course textbook (Rathus, 2005) and controlled the level of difficulty.

Participants in the third semester of data collection (n = 43) also completed a critical thinking exercise before and after the demonstration. Students evaluated a print advertisement selling a weight loss product listing strengths, weaknesses, validity indicators, and threats to validity. Students rated the plausibility of the rationale for the product on a 7-point Likert scale ranging from 0 (not at all plausible) to 6 (very plausible) and the likelihood of purchasing the product on a scale from 0% to 100% to evaluate potential behavior change. After the class demonstration, students evaluated a second weight loss advertisement. Students then rated the extent to which they enjoyed the class activity from 1 (not at all enjoyed) to 7 (very much enjoyed) and the validity of the task for teaching the topic from 1 (not at all valid) to 7 (very valid).

The classroom demonstration involved an experimental comparison between two sodas in which students indicated their taste preference. It occurred during the second week of classes in a 75-min class period as part of the research methods topic. I began by announcing that I enjoyed the taste of a particular brand-name soda and served as a consultant for that soda company. I described my intent to gather data regarding taste preference to replace the existing soda machines on campus with those of the rival, preferred soda. Volunteers tasted the regular version of the rival soda and a flavored (e.g., lime) diet version of the campus soda, indicating their taste preference. One student volunteer tallied preference data. Students identified procedures that either supported or detracted from the validity of the experiment, identified the type of validity demonstrated, and suggested alternate means of increasing validity. During the experiment, I committed multiple errors to increase the likelihood that students identified the rival soda as the preferred soda, threatening internal and external validity (see summary in Table 1). I ended the experiment when the rival soda was ahead and concluded that all students preferred the rival soda and thus soda machines across university campuses should be replaced with the rival cola. Following the demonstration, students discussed supports and threats to validity, implications of validity threats, and means to improve the experiment.

Results

A mixed ANOVA, with time of assessment as the within-subjects factor and class as the between-subject factor, evaluated the impact of the demonstration on comprehension of validity. Results indicated that students significantly increased the percentage of questions answered correctly across the quizzes, F(2, 71) = 47.24, p < .001. The effect size was .58. Pairwise comparisons indicated that each of the three assessment points differed significantly from each other. On average students correctly answered 39.7% of questions at the beginning of the semester (95% confidence interval = 30.5–48.9%), 67.3% before the demonstration (95% confidence interval = 61.3–75.2%), and 89.0% following the demonstration (95% confidence interval = 82.2–95.9%). The between-subject factor
Table 1. Summary of Demonstrated Compromises to Validity

<table>
<thead>
<tr>
<th>Demonstrated Validity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threats to internal validity</td>
<td>Announce consultant role to rival soda company</td>
</tr>
<tr>
<td></td>
<td>Announce intent of demonstrated experiment</td>
</tr>
<tr>
<td></td>
<td>Announce instructor’s rival soda preference</td>
</tr>
<tr>
<td></td>
<td>Vary sodas by calorie count, caffeine, and flavor</td>
</tr>
<tr>
<td></td>
<td>Call on some participants to participate</td>
</tr>
<tr>
<td></td>
<td>Goad participants (gently) who first decline to volunteer</td>
</tr>
<tr>
<td></td>
<td>Present sodas without counterbalancing or blinding</td>
</tr>
<tr>
<td></td>
<td>Present rival soda first</td>
</tr>
<tr>
<td></td>
<td>Compliment rival soda on presentation (e.g., “Here’s the good one”)</td>
</tr>
<tr>
<td></td>
<td>Offer no palate cleanser (e.g., crackers)</td>
</tr>
<tr>
<td></td>
<td>Question student preference of campus soda (e.g., “Are you sure about that?”)</td>
</tr>
<tr>
<td></td>
<td>Compliment student preference of rival soda (e.g., “I thought so”)</td>
</tr>
<tr>
<td></td>
<td>Comment on running results for rival soda (e.g., “The better one is winning”)</td>
</tr>
<tr>
<td></td>
<td>Stop experiment when rival soda is in the lead by checking data tallies often</td>
</tr>
<tr>
<td>Threats to external validity</td>
<td>Choose volunteers of only one gender and ethnicity</td>
</tr>
<tr>
<td></td>
<td>Choose a relatively small sample</td>
</tr>
<tr>
<td></td>
<td>Conclude that all students prefer rival soda</td>
</tr>
<tr>
<td></td>
<td>Conclude that rival soda is best of any soda</td>
</tr>
<tr>
<td></td>
<td>Conclude all campuses of university should switch to rival soda</td>
</tr>
<tr>
<td>Support of validity</td>
<td>Collect data</td>
</tr>
<tr>
<td></td>
<td>Present two sodas</td>
</tr>
<tr>
<td></td>
<td>Tally data</td>
</tr>
<tr>
<td></td>
<td>Report results of data collection</td>
</tr>
<tr>
<td></td>
<td>Use student participants at university targeted for rival soda</td>
</tr>
</tbody>
</table>

was not significant, indicating that all three classes improved across the three assessment points. A t test comparing the percentage of correctly answered exam questions indicated that students performed significantly better on the nine validity questions (M = 74%, SD = 14.1) than the remaining 51 items (M = 64%, SD = 12.6), t(73) = 36.10, p < .001. The effect size was .69.

A series of six repeated measures ANOVAs, with time serving as the within-subjects factor, analyzed the total correct responses for each of the four open-ended validity questions, and the average score for likelihood and plausibility rankings, to evaluate the critical thinking exercise. Results indicated that students significantly increased the number of correctly identified weaknesses, F(1, 40) = 13.89, p < .005; validity indicators, F(1, 39) = 9.82, p < .005; and threats to validity, F(1, 39) = 22.03, p < .001 (see Table 2). The effect sizes were .26, .20, and .29, respectively. The trend for students to reduce the likelihood of purchasing the product approached significance, F(1, 39) = 3.84, p = .057, with an effect size of .09. Students did not differ in the number of correctly identified strengths or plausibility rankings (all ps > .05). Finally, a majority of students (82%) indicated they enjoyed the task (M = 4.6, SD = 1.1), and 86% indicated it was a valid classroom demonstration to learn about validity (M = 4.7, SD = 1.4).

Discussion

Results supported the use of a classroom demonstration of an experiment with compromises to validity as a means to increase comprehension of the scientific method. Students increased quiz performance significantly after the demonstration, as compared to quiz scores before the demonstration and at the beginning of the course. Two baseline assessments controlled for knowledge about validity that occurred prior to the course (Assessment 1), as well as knowledge obtained from reading the textbook (Assessment 2).

Table 2. Means and Standard Deviations for Critical Thinking Assessment Evaluating a Print Advertisement

<table>
<thead>
<tr>
<th>Validity Indicator</th>
<th>Predemonstration</th>
<th>Postdemonstration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Strengths</td>
<td>2.02</td>
<td>0.93</td>
</tr>
<tr>
<td>Weaknesses</td>
<td>1.62</td>
<td>0.73</td>
</tr>
<tr>
<td>Validity indicators</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Threats to validity</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td>Plausibility</td>
<td>2.78</td>
<td>1.21</td>
</tr>
<tr>
<td>Self purchase</td>
<td>19.17</td>
<td>27.80</td>
</tr>
</tbody>
</table>

*p < .005. †p = .057.
Students also correctly answered a greater percentage of exam items on experimental validity compared to other topics taught by more traditional methods such as lecture and discussion. This finding suggests that active learning, in which students directly encounter the phenomenon under study, may increase interest and comprehension of the topic at hand (Kolb, 1984). One possible mechanism to account for this finding is that the active learning demonstration provided subjective personal meaning about the concept of validity and a reference for testing ideas about validity, predictors of increased research interest (Vittergol et al., 2004). This finding suggests that active learning may be beneficial in teaching alternate areas of psychology as well.

The demonstration of instructor ineptitude aimed to enhance students' critical thinking skills and to transfer the use of these skills to out-of-class situations. The ability to think through an issue is an essential outcome of education (McCarthy, 2005). This exercise encouraged this student outcome and emphasized at an introductory level that psychology provides an approach to problems in their lives. The critical thinking assessment indicated that students significantly increased the number of listed validity indicators and threats in a novel task, suggesting that generalization occurred. Examples of listed compromises to validity included reliance on testimonials, lack of controlled conditions, and small sample sizes. Also, students did not merely reiterate terminology, but rather responded in their own words (e.g., “Has pictures of people that used the product before & after”). The ease of identifying threats to validity may be due to the fact that I chose advertisements with questionable validity to highlight the need to use critical thinking skills in daily life. Students reported examples of valid research reports aimed at the general public, such as medication claims for insomnia, in the ensuing discussion, which implies that generalization occurred. Examples of listed compromises to validity included reliance on testimonials, lack of controlled conditions, and small sample sizes. Also, students did not merely reiterate terminology, but rather responded in their own words (e.g., “Has pictures of people that used the product before & after”). The ease of identifying threats to validity may be due to the fact that I chose advertisements with questionable validity to highlight the need to use critical thinking skills in daily life. Students reported examples of valid research reports aimed at the general public, such as medication claims for insomnia, in the ensuing discussion, which implies that the demonstration did not increase student cynicism about research in general.

The demonstration also provided a lively atmosphere that encouraged students' critical thinking. Not only did students list errors of validity, but they also discussed the implications these errors created and designed alternate procedures to address the concern. Students consistently identified blatant errors in validity, such as the use of two sodas that differed by more than name brand. They also noted subtleties in this manipulation, such as the perception that fewer men chose the diet soda because they are less likely to diet and that students did not cleanse their palate between tasting sodas. The procedure was also engaging, as students reported they were trying to catch their instructor making “slips,” and I reminded them several times to refrain from commenting on the procedures until the discussion. Students quickly pointed out the selection bias in volunteers and engaged in conversations with each other (rather than mediated through the instructor) about the implications that results did not reflect many unique demographics, leading to discussions of external validity.

Several limitations exist in the evaluation of the efficacy of this demonstration. First, I could not compare the unique learning benefits of the demonstration versus a lecture-oriented class because of the absence of random assignment to class with or without the demonstration. Second, the postdemonstration assessment of a second weight loss advertisement provided a weaker demonstration of transfer than would an advertisement for an alternate product. Finally, the evaluation component did not include a specific indicator of student engagement. Future use of this demonstration might include student planning of classroom experiments with errors of validity to increase student participation in the planning of experiments and the experiential process. Overall, the demonstration provided an interactive and entertaining means of demonstrating scientific concepts to students in a general psychology class.

References


**Notes**

1. I thank George Allen for his thoughtful comments on an earlier draft of this article.
2. Send correspondence to Kimberli R. H. Treadwell, Department of Psychology, University of Connecticut, 406 Babbidge Road Unit 1020, Storrs, CT 06269–1020; e-mail: kimberli.treadwell@uconn.edu.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
Griggs and Zechmeister (2000; M = 691.1). However, if you extrapolate terms per chapter, this difference of about 85 terms is not much given the number of chapters in the texts (M = 24.3 for our sample and M = 17.9 for the Zechmeister & Zechmeister, 2000, sample). Therefore, the difference amounts to only 3 to 5 terms per chapter regardless of which mean is used for the calculation.

After eliminating duplicate terms, 1,287 terms remained. As observed by Zechmeister and Zechmeister (2000), most were unique to one text. There were 1,048 (75.7%) unique terms, 246 (17.6%) terms appearing in two glossaries, and only 93 terms (6.9%) appearing in all three glossaries. Thus, as with contemporary introductory textbooks, no truly substantial common core vocabulary existed in these 1950s textbooks. Given that we chose texts from the 1950s to maximize the likelihood of finding a substantial common core, our failure to find one indicates that such a core has likely never existed.

We did, however, find 109 classic core vocabulary terms. These terms appear in Table 1. To facilitate their use by introductory psychology teachers, we grouped them in accordance with the chapter topics and organization of contemporary introductory psychology textbooks. These terms comprise about one third of the core vocabularies observed for both the 1950s and 1990s textbooks.

We next considered how these classic terms were distributed across the major topic areas within introductory psychology. As we expected, few were in areas that have grown and gained prominence since the 1950s, especially cognitive psychology. Only 4 (3.7%) of the 109 terms are cognitive in nature. Most of the terms come from older research areas—sensation/perception (16; 14.7%), learning (16; 14.7%), and physiological psychology (18; 16.5%). This latter percentage would be even greater if we had included the developmental psychology terms, which all concerned genetics or physical development, in the physiological category. It is also interesting to note that both the personality and therapies topics were dominated by Freudian or psychoanalytic terms (5 out of 9 and 5 out of 8, respectively).

In conclusion, it appears that a substantial core vocabulary in introductory textbooks does not presently exist and probably has never existed. However, a classic core vocabulary of over 100 terms does exist. As Zechmeister and Zechmeister (2000) pointed out, the American Psychological Association’s Committee on Undergraduate Education (McGovern, Furumoto, Halpern, Kimble, & McKeachie, 1991) recommended that psychology teachers use the principle that “less is more” in their courses. In response, Zechmeister and Zechmeister (2000) asked with respect to the introductory course, “If less is more, then what should that less be?” (p. 7). Although we agree with Zechmeister and Zechmeister that exactly what that “less” should be is unclear, we believe that the classic core vocabulary should definitely be a part of it.

References


Notes

1. We thank Jeanne Zechmeister for providing us with the core concept data from Zechmeister and Zechmeister (2000), three anonymous reviewers, and Randolph Smith for valuable comments on an earlier version of this article.

2. Send correspondence to Richard A. Griggs, Department of Psychology, PO Box 112250, University of Florida, Gainesville, FL 32611; e-mail: rgriggs@ufl.edu.

Classroom Demonstrations of Auditory Perception

LaDawn Haws
Department of Mathematics and Statistics
California State University, Chico

Brian J. Oppy
California State University, Chico

Many faculty who teach a psychology class with a sensation and perception component present a variety of demonstrations of visual perception, but few about audition. Demonstrations using inexpensive materials can illustrate some of the basic concepts related to auditory perception. In this article we describe simple and inexpensive demonstrations of sound localization, wave cancellation, frequency/pitch variation, and the influence of different media on sound propagation.

Vol. 29, No. 2, 2002 147
Faculty who teach courses in sensation and perception and classes in introductory perception psychology commonly present demonstrations of visual perception phenomena (Goldstein, 1999). However, they seldom present demonstrations of auditory perceptual phenomena. In this article we present several simple inexpensive activities involving audition that give students concrete experiences to complement the abstract concepts presented in texts.

Localization Demonstration 1

Each pair of students needs a 30-in. length of 3/8-in. outer diameter flexible plastic hose, available from any hardware store. The exact length is not important, but the midpoint should be marked clearly and accurately. The Listener holds the hose with one end at each ear, eyes closed. The Tapper taps the hose gently with a pencil, and the Listener must determine if the tap was closer to the left ear or the right ear. The Tapper should tap at different distances, ranging from 12 in. to 1 in. away from the center mark. The Tapper must be careful to not tap too vigorously, which would give the Listener a tactile directional clue. After several trials, the Listener and Tapper exchange roles.

This demonstration gives students an appreciation of the accuracy with which people can determine the location of a sound source. Most people are able to distinguish the tapping direction even when the tap is only 1 in. away from the center mark of the hose. Listeners locate sound sources largely due to the interaural time difference (Goldstein, 1999), defined as the minute differences in time for the sound waves to reach one ear before the other. The sound source is closer to the ear that received the input first. Interaural intensity differences also provide a significant cue to location: The stimulus is slightly more intense (louder) for the closer ear.

Localization Demonstration 2

This demonstration is done in two parts and requires a "clicker" (the type used by children on Halloween works well) and a 12-in. length of 3-in. diameter plastic pipe or a heavy cardboard tube. For Part 1, the Listener sits facing the class, eyes closed. The Clicker stands behind the Listener and makes a "click" in the same plane as the listener, parallel to the front wall. The Listener must point to the direction of the click. The Clicker clicks 2 to 3 ft from the Listener, in several positions including left, right, and directly overhead, but not in front or back. The Listener will be very accurate in detecting the direction of the click.

For Part 2, the Listener holds the pipe firmly up to one ear and the Clicker repeats the demonstration. The Listener is likely to perceive an overhead click to be closer to the ear without the pipe because the sound waves must travel farther to reach the "pipe ear" due to the interaural time difference. This demonstration illustrates two basic points. First, Listeners are quite accurate in locating a sound stimulus in three-dimensional space. Second, sound localization is very "bottom up"; even though Listeners are aware of the tube, they are quite confident of their impression of direction—a misplaced confidence when the tube is in place.

Frequency and Pitch

The Demonstrator pours water into a 4-ft length of 1-in. diameter metal pipe with the bottom end plugged. As the water fills the pipe, the vibrations generated by the moving water produce an audible roar. The pitch of the roar will increase as the pipe fills. Once the pipe is full, the Demonstrator empties the pipe and repeats the demonstration, but this time while the Demonstrator pours the water, an assistant taps the lower part of the pipe continuously. The pitch of the tapped pipe will get lower as the pipe fills. This demonstration works well even in a large auditorium—simply hold a microphone near the top opening of the pipe.

The two versions of this demonstration produce opposite effects because different media are vibrating. In both cases, a shorter span of media produces higher frequency sound waves. In the first demonstration, the vibrating column of air becomes shorter as the pipe fills, and the increasingly shorter column produces an increasingly higher frequency (and higher pitch). In the second demonstration, the tapping causes the water to vibrate. The column of water grows longer as the pipe fills, producing an increasingly lower pitched sound. Of course there are other demonstrations of the relationship between length and pitch (e.g., wind chimes and xylophones) but this experiment is interesting because there are two columns changing length and pitch continuously in opposite directions.

Sound Cancellation

When two waves of the same frequency and same phase combine, the resulting wave retains the common frequency, but the amplitude of the resultant wave is the sum of the amplitudes of the combining waves, as in Figure 1. People perceive the increased amplitude of sound waves as a louder sound. Conversely, if two waves of the same frequency but half a wavelength out of phase combine, the two waves annihilate each other—the amplitude is zero, as in Figure 2. In theory, two sound waves of the same pitch but half a wavelength out of phase should combine to produce silence. In practice, it is difficult to achieve total cancellation, but this demonstration makes it possible to verify that cancellation does indeed occur.
The equipment required for this demonstration is a stethoscope that has been dismantled and refitted with Y connectors and rubber hoses as in Figure 3; a 440-A tuning fork is also required. The upper hose on the modified stethoscope is 1.25-ft long between the Y connectors, and the bottom hose is 2.5 ft. The lengths of the hoses are related to the frequency of the tuning fork and should be measured carefully.

The Listener dons the modified stethoscope, putting the earpieces into both ears. The Tapper raps a 440-A tuning fork and holds it close to (but not touching) the open end of the hose. When the sound waves reach the first Y connector, roughly half will enter the top hose and half the bottom hose (remember, the bottom hose is one full wave length, and the top hose is half a wavelength). The sound waves traveling along the bottom hose will reach the second Y connector after completing one full cycle—the waves will be starting the positive arch as they enter the Y. The sound waves traveling along the top hose will reach the second Y connector after completing half a cycle—the waves will be starting the negative arch as they enter the Y. The two kinds of waves entering the Y look just like the waves in Figure 2. The Listener hears a very soft sound because many of the waves cancel as they meet at the Y.

If the Listener pinches one of the tubes, the volume of the sound will increase! Most listeners will be surprised at the counterintuitive increase in volume that results from cutting off one of the hoses. When both hoses are open, sound cancellation occurs. When the Listener pinches one of the hoses, the sound waves travel directly down the open hose with no cancellation—that is why the sound is louder.

The phenomenon of cancellation has a practical application: A jackhammer operator wearing a set of earphones that produce waves that cancel the jackhammer waves can greatly reduce the damage done to his or her inner ears. A physics text (Halliday, Resnick, & Walker, 1996) gave a rigorous exposition of the physics behind sound phenomena; a discussion aimed at readers with less technical background is also available (Hewitt, 1998).

Sound Traveling Through Different Media

The apparatus for this demonstration is a metal clothes hanger suspended from two 20-in. lengths of string (exact length is not important) as in Figure 4. The Listener wraps one string around each index finger, then sticks those fingers into his or her ears, letting the hanger dangle loosely. The Tapper strikes the hanger with a pen. The Listener will hear what sounds like church bells ringing in his or her ears, but a nearby observer hears nothing more than a dull thud.

Sound waves are affected by the different media through which they travel. In this case, sound waves traveling through air dissipate much faster than waves traveling through bones and string, so the Listener receives virtually all of the tones and overtones that are produced when the hanger is struck by the pen, whereas a nearby observer receives only the dissipated tones. Lowery (1997) described this and other activities.

Conclusions

We have used these simple activities with learners from different disciplines and at different age levels. Although we have done no formal study to evaluate the pedagogical effectiveness...
of these activities, students’ efforts on homework problems indicate that learning has indeed occurred—students use language correctly and explain concepts in their own words. Perhaps most important, in course evaluations, students comment on how much they enjoyed the “toys” and that they had learned so much because of the hands-on approach.

References


Send correspondence to LaDawn Haws, Department of Mathematics and Statistics, California State University, Chico, CA 95929–0525. E-mail: lhaws@csuchico.edu.

Schema Theory: A New Twist Using Duplo™ Models

Joe D. Nichols
School of Education
Indiana University–Purdue University Fort Wayne

I describe a classroom demonstration for teaching and discussing Piaget’s concepts of schemata, assimilation, accommodation, and equilibration. One student (“teacher”) describes an asymmetrical construction built from children’s Duplo™ blocks such that another student (“learner”) can build an identical model. The teacher and learner cannot see each other, and the learner may not speak. The teacher must use verbal instructions only to guide the construction of the model. Because the students are seated back to back, the learner cannot see and the teacher cannot see how well the learner is following the instructions. At its completion, the learner’s structure may resemble the model but the two can be quite different. After this demonstration, students analyze how Piaget’s theory relates to this situation.

Observations

Before class, I construct an asymmetrical structure using 25 to 30 components made of children’s Duplo™ blocks. These blocks are in various shapes and colors of hard plastic, larger than LEGO®, to facilitate every class member’s view of the demonstration. I bring the construction in a closed box and bring in a separate bag of 25 to 30 identical components unassembled. Two student volunteers sit back to back at desks in front of the class. One student (designated the teacher) receives the preassembled construction and the other student (designated the learner) receives the matching unassembled pieces. The task is for the teacher to describe the model and give instructions so that the learner can build a matching model. The learner is not allowed to speak. The teacher must use verbal instructions only to guide the construction of the model. Because the students are seated back to back, the learner cannot see and the teacher cannot see how well the learner is following the instructions. At its completion, the learner’s structure may resemble the model but the two can be quite different. After this demonstration, students analyze how Piaget’s theory relates to this situation.
Seeing the Light: A Classroom-Sized Pinhole Camera Demonstration for Teaching Vision

Matthew W. Prull
Whitman College

William P. Banks
Pomona College

We describe a classroom-sized pinhole camera demonstration (camera obscura) designed to enhance students' learning of the visual system. The demonstration consists of a suspended rear-projection screen onto which the outside environment projects images through a small hole in a classroom window. Students can observe these images in a darkened classroom. Instructors can demonstrate the function of the lens and pupil and the structural basis for nearsightedness and farsightedness. Students who saw the demonstration as part of a lecture on the visual system learned more (i.e., showed greater performance gains from pretest to posttest) than a comparable group of students who received the lecture only. Students reacted favorably to the demonstration. These data suggest that incorporating the demonstration into class presentations on vision can improve student learning.

Sensation and perception are important but challenging topics in introductory psychology. Content analyses of introductory texts suggest that sensation and perception chapters are among the most lengthy (Griggs, Jackson, Christopher, & Marek, 1999), contain the greatest number of psychological terms and concepts (Landrum, 1993), and are among the 10 most frequently assigned chapters (Miller & Gentile, 1998). Many concepts in sensation and perception do not come easily to students, but well-chosen pedagogical aids could ease the difficulty of learning this material.

Especially prominent in the introductory chapters on sensation is the information about vision and visual processes. Indeed, most introductory texts include information about the structure of the mammalian eye, which implies that knowledge of the biological basis of vision is important in introductory psychology. Although instructors who cover vision in their courses can use several demonstrations such as visual illusions (Dougherty, 1990; Klopfer & Doherty, 1992), color vision (Beins, 1983; Horner, 1997), Gestalt principles of organization (Kozub, 1991), and adaptation to displaced vision (Terborg, 1990) to enhance learning of visual processes, we have found relatively few demonstrations that aid teaching about the structure of the visual system.

This article describes a pinhole camera demonstration to enhance teaching the structure and function of the eye. Photography classes frequently employ a pinhole camera, or camera obscura, to aid learning about the workings of the camera and the nature of light (Pirenne, 1970). The pinhole camera can also illustrate some of the functions of the eye (Wade & Finger, 2001). The traditional pinhole camera consists of a box with a small hole at one end. Light enters the pinhole (the “pupil”) and casts an inverted image of the external en-
vironment on the inside box end opposite the pinhole (the “retina”). The size of the pinhole determines image brightness and focus, with larger hole diameters producing brighter but blurrier images. Although pinhole cameras are typically small in scale, our room-sized version of the pinhole camera allows us to teach a classroom of students about the structure of the visual system.

Method

Participants

Sixty-two students provided data on the learning effectiveness of the demonstration (the learning effectiveness group). Sixty-seven additional students provided their subjective reactions to the demonstration (the self-report group). All were traditional-age students at Whitman College (ages 18 to 22) who enrolled in the first author’s introductory classes between 2001 and 2003. All students read the same textbook and followed the same order of topics.

Materials

We used a lightweight rear-projection screen, an inexpensive glass lens, and a large piece of heavy black fabric purchased at a local fabric store. The screen was a custom-made 29 × 38 in. lightweight portable “minifold” screen (#40466C, Da-Lite Screen Company, Warsaw, IN), although low-cost alternatives can yield equally effective results. For example, instructors can manufacture an effective screen by mounting a large sheet of vellum, waxed paper, translucent plastic (e.g., that used in some kinds of trash bags), or frosted Plexiglas over a lightweight wooden frame. The lens was a 75-mm diameter plano-convex glass lens with a focal length of 400 mm (#J53103, Edmund Industrial Optics, Barrington, NJ). The fabric was 1 ft larger than the dimensions of one of the classroom windows. We cut three holes of different diameters (approximately ½, 1, and 2 in., respectively; each separated by about 2 in.) into the fabric at a height that was easily accessible to the instructor once we mounted the fabric over the window. These holes served as the pinholes in the demonstration. We covered each hole with small square flaps made from the same material, and we fastened the prepared fabric to the wall surrounding the window using strapping tape so that the window was completely covered. Once the fabric was in place, we hung the screen from the ceiling using heavy cord and hooks, directly in front of the three holes in the fabric. The distance between the screen and the fabric was about 2 ft (see Figure 1). The blinds and curtains on all other classroom windows were closed to make the room as dark as possible.

Design and Procedure

For the learning effectiveness group, the design was a 2 (test occasion: pretest, posttest) × 2 (condition: experimental, control) mixed factorial. The experimental condition (n = 38) received the pinhole camera demonstration but the control condition (n = 24) did not. Experimental and control conditions came from different class sections.

Students in the learning effectiveness group took the 10-item quiz the day before the lecture to assess baseline knowledge. The goals of the following day’s 50-min lecture were to increase students’ understanding of (a) the structure and function of various components of the eye, (b) the organization of visual pathways leading from eye to brain, and (c) receptive field characteristics in neurons of the primary visual cortex. The aim of the demonstration was to enhance understanding of the structure and function of the eye. For the experimental condition, the demonstration began immediately after introducing the day’s topic. In the control condition,
the instructor replaced the demonstration with lecture material that covered the same information.

Students in the experimental condition gathered around the side of the screen that faced the classroom. The instructor then asked an assistant to turn out the lights. Following preliminary comments about the light-collecting nature of the eye, the instructor explained the function of the pupil by initially lifting the flap that covered the ½ in. diameter hole on the fabric. Opening the flap permitted a small amount of light to enter the room and fall on the side of the screen facing the window; the light projected an inverted image of the outside environment that students could easily view on their side. The instructor mentioned that the inverted screen image was similar to the inverted retinal image.

The first image was fairly dim but relatively focused, but by closing the ½ in. hole and opening the larger holes one at a time the instructor illustrated the trade-off between image brightness and image clarity with various “pupils” (larger holes produce a brighter but blurrier screen image; Wald, 1950). When students understood that an ideal image for the visual system would be both bright and focused, the instructor brought out the lens and held it between the open largest diameter hole and the screen. The result was a bright, highly detailed inverted image of the outside environment.

Next, the instructor introduced the structural basis of nearsightedness and farsightedness. Nearsightedness (myopia) can arise when the retina is too far from the focal point of the lens. The instructor simulated nearsightedness by holding the lens steady with one hand and moving the screen away from the window (i.e., toward the students) with the other hand. This procedure caused the focused image to blur. In contrast, farsightedness (hypermetropia) can emerge when the distance between lens and retina is too short; the lens essentially focuses the image at a point behind the retina. Moving the screen closer to the lens simulates farsightedness, and again the once-focused image blurs. Moving the screen thus illustrated the precise distance relation between lens and retina that is necessary for a clear retinal image (instructors may wish to note, however, that nearsightedness and farsightedness should not be equated with general blurriness). In total, the demonstration took about 15 min. The remaining class time consisted of a more traditional lecture designed to meet the remaining teaching goals. One week after this class meeting, the instructor administered the same 10-item quiz as a posttest.

All students in the self-report group experienced the demonstration but responded to the pretest questionnaire at the beginning of the class session and took the posttest questionnaire at the end of that class period.

Results

Learning Effectiveness

We conducted a 2 (condition: experimental, control) × 2 (test occasion: pretest, posttest) mixed ANOVA on the quiz scores using α = .05. Students in the experimental condition had higher mean scores than students in the control condition (M = 5.78 vs. 5.10), F(1, 60) = 4.76, p < .05, partial η² = .07, and posttest scores were higher than pretest scores (M = 6.39 vs. 4.65), F(1, 60) = 58.62, p < .05, partial η² = .49. Most important, condition and test type interacted, F(1, 60) = 7.35, p < .05, partial η² = .11. Further analyses suggested that pretest scores did not differ between the experimental and control conditions (M = 4.68 vs. 4.58, respectively), F(1, 60) = 0.10, p > .05, partial η² = .002, but posttest scores were reliably higher in the experimental condition than in the control condition (M = 6.87 vs. 5.63, respectively), F(1, 60) = 8.98, p < .05, partial η² = .13.

Self-Reports

Mean posttest values for each of the three items were reliably higher than mean pretest values, indicating an increased sense of learning. Respective pretest and posttest means were 4.08 and 5.55 (SD = 1.54 and 0.84, respectively) for Item 1, F(1, 66) = 76.19, p < .05, partial η² = .54; 4.52 and 6.02 (SD = 1.46 and 0.73) for Item 2, F(1, 65) = 60.00, p < .05, partial η² = .48; and 4.42 and 6.24 (SD = 1.70 and 0.72) for Item 3, F(1, 66) = 69.87, p < .05, partial η² = .51. (Error df are 65 in the analysis for Item 2 because 1 student did not respond to that item.)

As shown in Table 1, responses to each of the six additional posttest questionnaire items indicated positive reactions. Open-ended written comments were also quite positive: “That was awesome! I have never seen anything like that.” “Big help for me to understand the relationships between vocab and actual happenings,” and “I have taken extensive photography courses and that was by far the best pinhole camera I have ever seen.”

![Table 1. Means and Standard Deviations for the Self-Report Group’s Posttest Questionnaire Items](image)

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The pinhole camera demonstration enhanced the instructor’s lecture on vision</td>
<td>6.67</td>
<td>0.61</td>
</tr>
<tr>
<td>The pinhole camera demonstration enhanced my understanding of the structure and function of various components of the eye</td>
<td>6.36</td>
<td>0.83</td>
</tr>
<tr>
<td>Relative to lecturing on the topic, the pinhole camera demonstration was more effective in its ability to communicate to students the structure and function of various components of the eye</td>
<td>6.19</td>
<td>0.86</td>
</tr>
<tr>
<td>The pinhole camera demonstration did not help me understand better how nearsightedness and farsightedness relate to the structure of the eye</td>
<td>1.45</td>
<td>0.69</td>
</tr>
<tr>
<td>The pinhole camera demonstration was a waste of class time</td>
<td>1.23</td>
<td>0.64</td>
</tr>
<tr>
<td>I recommend that the pinhole camera demonstration be included in future introductory psychology courses</td>
<td>6.42</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Note. Based on a scale ranging from 1 (strongly disagree) to 7 (strongly agree).
Discussion

These results suggest that a room-sized pinhole camera demonstration enhances learning about vision. Not only did students react favorably to the demonstration, but students who saw the demonstration as part of a lecture on vision learned more (as measured by posttest quiz scores) than a group of students who did not see the demonstration.

The effectiveness of this demonstration in classroom settings depends on at least five elements. First, the screen must be fairly large so that the image is large enough for the entire class to see. A screen size of 3 ft $\times$ 3 ft works well for class sizes of up to 30 students. Larger classes will require larger screens. Second, the screen should be suspended from the ceiling or mounted on a stand so that both of the instructor’s hands remain free. Manipulating both lens and screen is cumbersome. Third, the classroom must have windows as well as blinds or curtains so that the room can be made as dark as possible (instructors who teach in rooms without windows will need to schedule their class in such a location to implement this demonstration). Fourth, the lens must have a long focal length (400 mm works well) because a shorter focal length will yield a clear image that is too small for the class to see. Fifth, the instructor must have time for room preparation prior to the class session.

Instructors can use the demonstration in sensation, perception, and cognitive courses, as well as introductory courses, and can implement the demonstration in lecture or in a laboratory assignment in which students manipulate the lens and screen themselves. Instructors may even assign the construction of a classroom-sized or traditional-sized pinhole camera as a class project. Regardless of how instructors choose to use it, we conclude that the pinhole camera demonstration is well received and empirically validated as an effective teaching aid on the visual system.

References


Notes

1. We thank Melissa Clearfield, Katalina Diamond, and Kathleen Felton, in whose classes we demonstrated the pinhole camera. We also thank Deborah DUNANN Winter for editorial suggestions.

2. Send correspondence to Matthew W. Prull, Department of Psychology, 345 Boyer Avenue, Walla Walla, WA 99362; e-mail: prullmw@whitman.edu.
Results

A 2 x 2 mixed model ANOVA revealed significant main effects for both experimental condition, $F(1, 58) = 17.90, p < .01, \eta^2 = .29$, and assessment time, $F(1, 58) = 7.03, p < .01, \eta^2 = .12$. At the initial assessment, students in the self-generated mnemonic condition had a greater percentage of recall ($M = 85.20$) than their counterparts ($M = 73.83$). This trend continued during the second assessment with self-generated mnemonics leading to a higher percentage of recall ($M = 84.10$) than provided mnemonics ($M = 67.60$). The analysis also revealed a statistically significant interaction between experimental condition and time of assessment, $F(1, 58) = 3.53, p < .05, \eta^2 = .06$. Students in the provided mnemonic condition showed a deterioration of recall at the second assessment, but students in the self-generated mnemonic condition did not.

Discussion

Researchers may debate the usefulness of first-letter mnemonics, but for students their value is well established (Carlson et al., 1981). If students intend to use such devices to aid themselves in learning the cranial nerves, then physiological psychology instructors can maximize the likelihood of their success. Results of this study indicate that student success is facilitated by having students take an active role in the creation of the mnemonic device they use. Students who generated their own mnemonic devices had a greater percentage of recall than their counterparts at both assessment times. Perhaps the most important result regards the interaction between mnemonic condition and assessment time. Whereas students assigned to the provided condition showed a deterioration of recall at the second assessment, students in the self-generated condition did not. The failure to find a statistically significant decrease over time for the self-generated condition indicates an important advantage to this method. The self-generated mnemonics appeared to provide longer lasting success in the ability of students to recall the cranial nerves.

One could argue that students in the self-generated mnemonic condition performed better simply because they may have spent more time with the material than students in the provided mnemonic condition. A control group instructed to learn the cranial nerves with the use of a nonmnemonic method, although excluded, would have been helpful in addressing this question. Nonetheless, the study does show that, given the same amount of time to learn, the students using their own mnemonic rather than using the commonly provided one displayed better learning of the cranial nerves.

Courses in physiological psychology include a number of important, difficult terms and concepts. Although it is impossible for any professor to “make” a student learn, professors can maximize the likelihood of student learning. The results of this study point to an additional aid that instructors can emphasize to help students retain these difficult ideas. More important, these results demonstrate another example of the power of student engagement to bring about student learning. By encouraging students to take a more active role in the material they are attempting to learn, instructors maximize the likelihood of their students’ success.

References


Note

Send correspondence to Christopher M. Bloom, Department of Psychology, University of Southern Indiana, 8600 University Boulevard, Evansville, IN 47712; e-mail: cbloom@usi.edu.


W. Robert Batsell, Jr.

Kalamazoo College

Cogan and Cogan (1984) introduced a classroom demonstration of classical conditioning that involved pairing a neutral cue with lemonade powder. This article reports an addition to the Cogan and Cogan method to incorporate the phenomenon of renewal. Renewal, a relatively new phenomenon in classical conditioning, occurs when acquisition and extinction occur in different contexts. If testing occurs in the acquisition context or a neutral context, a return of responding to the extinguished cue occurs. This outcome provides evidence that extinction is new, context-dependent learning. This exercise promotes understanding the procedures and results of renewal experiments as well as the theoretical and clinical implications of this phenomenon.

Cogan and Cogan (1984) described an effective demonstration of classical conditioning where students paired a neutral cue (Pavlov) with an unconditioned stimulus (US) of lemonade powder. Instructors could use this exercise to ex-
pose students to the phenomena of acquisition and extinction and, if desired during a subsequent class, the concepts of spontaneous recovery and reacquisition. This demonstration has retained its usefulness as evidenced by its inclusion in teaching handbooks (e.g., Ware & Johnson, 1996). Yet advances in extinction research have expanded knowledge of this phenomenon, and an update of the Cogan and Cogan exercise is in order.

Spontaneous recovery is the return of the conditioned response (CR) to an extinguished conditioned stimulus (CS) an extended period of time following extinction trials; the presence of spontaneous recovery indicates that extinction is not unlearning. Although spontaneous recovery of the CR indicates that some trace of the original CS–US association survives extinction, the phenomenon of renewal confirms that extinction represents new, context-dependent learning (e.g., Bouton & King, 1983). Renewal involves a three-stage procedure that can be demonstrated with two groups, AAA and ABA, which are differentiated by the contexts in which they receive acquisition, extinction, and renewal. Both groups receive CS–US pairings in the same context (A); next, AAA receives extinction trials in Context A whereas ABA receives extinction trials in Context B. After both groups have received an equivalent number of extinction trials to eliminate responding to the CS, both groups receive the CS in Context A. Although AAA shows little response to the CS, ABA shows a return (or renewal) of the CR. Thus, the CR exhibited by ABA during renewal shows that extinction is not unlearning of the original CS–US association; instead, this outcome suggests that extinction is new learning. Furthermore, the inclusion of other groups, such as ABC, confirms that extinction is context dependent. In ABC renewal, following acquisition and extinction, testing the CS in a neutral context (C) produces a return of the CR relative to the CR produced following AAA training.

The finding that extinction is new, context-dependent learning has implications for clinical treatment. For example, if a therapist conducts fear extinction technique (e.g., exposure therapy) solely in her or his office, it is conceivable that the patient would experience fear reduction only in the therapist’s office. Indeed, the patient may experience a relapse of fear every time he or she encounters the fear-producing CS in the real world (Bouton, 2002). The fact that extinction is new learning already has spurred changes in extinction therapy. Ressler et al. (2004) showed that the use of cognitive enhancers during virtual reality exposure therapy produced rapid reductions in fear extinction. Thus, psychology students should be familiar with the phenomenon of renewal for its theoretical and clinical implications.

Results and Discussion

I designed this exercise to demonstrate renewal in humans. I most recently used this exercise with 10 learning students who tested 25 participants (5 students included an ABC participant). Table 1 shows the percentage of AAA, ABA, and ABC participants reporting a salivation CR on selected acquisition (Trials 1, 5, and 10), extinction (Trials 5 and 10), and renewal (Trials 1 and 2) trials. The percentage of CRs was similar for all training conditions across acquisition and extinction; these data are similar to those reported by Cogan and Cogan (cf., 1984). The data of interest are from

Method

I designed this exercise to complement the Cogan and Cogan (1984) demonstration following consultation with the Kalamazoo College Institutional Review Board Chair. In this exercise, the students were participants in the original lemonade exercise in the classroom, but later acted as experimenters in the renewal exercise in their homes.

In the second week of a recent learning class, students participated in the acquisition and extinction phases of the lemonade conditioning exercise. Next, they received lecture information (i.e., extinction, spontaneous recovery, reacquisition, and renewal), take-home instructions, and three 0.23-oz Kool-Aid lemonade powder packets.

The instructions described the following procedures. Each (student) experimenter recruited two friends who were not in the class to serve as participants at the experimenter’s dorm room or apartment. The experimenter conditioned and tested each participant individually. After hearing a verbal description of the procedure, the participant gave his or her consent. Each participant received acquisition in a quiet room (Context A) with 10 interspersed test trials as described by Cogan and Cogan (1984); each test trial was preceded by 3 to 6 acquisition trials. On a test trial, if the experimenter presented the CS and said “test,” the participant did not apply the US. Then, the participant self-reported if he or she experienced salivation in response to the CS. Following the 10th acquisition test trial, the participant went to a different room (Context B). ABA participants received 10 extinction trials in this context whereas AAA participants immediately returned to the acquisition context (A) for extinction training. During extinction, each participant received 10 presentations of the CS alone, and self-reported his or her salivation CR on each trial. Following the 10th extinction test trial, each participant walked to an adjoining room, but ultimately, each participant returned to the original acquisition room (A) for renewal testing. Participants received two CS-alone presentations during renewal. Furthermore, I encouraged the students, but did not require them, to recruit a third participant for the ABC condition.

To assess student learning, the experimenters completed lab reports containing the method, results, and interpretation of the findings. Following the completion of the experiment, the students compiled their data for class discussion.

I designed this exercise to demonstrate renewal in humans. I most recently used this exercise with 10 learning students who tested 25 participants (5 students included an ABC participant). Table 1 shows the percentage of AAA, ABA, and ABC participants reporting a salivation CR on selected acquisition (Trials 1, 5, and 10), extinction (Trials 5 and 10), and renewal (Trials 1 and 2) trials. The percentage of CRs was similar for all training conditions across acquisition and extinction; these data are similar to those reported by Cogan and Cogan (cf., 1984). The data of interest are from
this update allows instructors to include the renewal effect. Therefore, instructors may need to make additional efforts to keep their students abreast of updates in classical conditioning.

I indirectly assessed student learning on numerous measures. First, from class discussion and the students’ papers, it was clear that the exercise helped students learn the design and results of renewal experiments. Second, all students correctly answered the renewal question on the first exam. Third, I assessed students’ long-term retention of this information on a cumulative final exam 8 weeks after they completed the exercise. A question addressed problems that might undermine the success of extinction therapy for phobias. All 10 students mentioned the context dependency of extinction, and 6 explicitly described the renewal effect. Fourth, at the end of the quarter, students evaluated the renewal exercise via a six-item questionnaire scored on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). The students’ responses were generally positive: Did the exercise help you understand the procedures of renewal experiments? (M = 4.4, SD = 0.7); Did the exercise help you understand that extinction is new learning, not unlearning? (M = 4.4, SD = 0.7); Did the exercise help you understand that extinction is context dependent? (M = 4.5, SD = 0.7); Did the exercise facilitate understanding of the clinical importance of problems if extinction is only conducted in a single context (i.e., therapist office)? (M = 4.4, SD = 0.7); Did the exercise help you on the first test? (M = 4.0, SD = 0.6); and Should this renewal assignment be used in future classes? (M = 4.3, SD = 0.6). It is evident the students agreed that the exercise was effective in advancing the procedures and implications of the renewal effect.

In conclusion, the lemonade conditioning exercise has been a reliable means of demonstrating acquisition and extinction of classical conditioning (Cogan & Cogan, 1984); this update allows instructors to include the renewal effect to expand students’ knowledge of the theoretical and clinical implications of extinction. Further, the indirect evidence suggests that the exercise was effective in promoting student learning. One final reason for instructors to consider this exercise is that although some current textbooks include information about the renewal effect and context dependency of extinction (Domjan, 2003), a number of recent learning texts (cf., Klein, 2002; Lieberman, 2004; Terry, 2003) or computer programs of associative learning (Alloway, Wilson, & Graham, 2005) fail to include the renewal effect. Therefore, instructors may need to make additional efforts to keep their students abreast of updates in classical conditioning.

I thank my classes of learning students for their participation in this project and Mark Bouton for his feedback on the article.

**References**


**Notes**

1. I thank my classes of learning students for their participation in this project and Mark Bouton for his feedback on the article.
2. Send correspondence to W. Robert Batsell, Department of Psychology, Kalamazoo College, 1200 Academy Street, Kalamazoo, MI 49006; e-mail: rbatsell@kzoo.edu.
Examining Memory Phenomena Through Flashbulb Memories

Mark Sudlow Hoyert
Indiana University Northwest

Cynthia D. O’Dell
Psychology Department and Women’s Studies Program
Indiana University Northwest

In this article, we describe a project examining flashbulb memories. Students enrolled in introductory psychology courses recorded their memories for the same event during the first class meeting of the semester and then again 2 months later during the memory section of the course. Students analyzed the data for content and consistency. The analyses served as a vehicle for introducing and demonstrating memory phenomena. The project improved student understanding of memory phenomena, and students reported that the project was valuable in learning about memory functions and observing memory errors and distortions.

Brown and Kulik (1977) reported that unexpected and emotionally charged events produce unusually clear and detailed memories. They gave these recollections the name “flashbulb” memories because it seemed that the event had been imprinted into memory with a level of detail, clarity, and durability reminiscent of photographs. They also suggested that flashbulb memories are the product of a unique memory system. However, more recent explorations suggest that they obey the same principles as more conventional memories (Heuer & Reisberg, 1990; Neisser, 1982). As such, examining flashbulb memories provides a unique and colorful opportunity for observing many basic memory processes, such as elaboration, reconstruction, schemas, perma-store, consolidation, emotion, and rehearsal, typically covered in an introductory psychology course.

This project is roughly patterned after a flashbulb memory study conducted by Neisser and Harsch (1992). In that study, the researchers asked participants to describe what they were doing when they first learned of the Challenger explosion and to rate their confidence in their memories. The researchers questioned their participants within 1 day of the disaster and again 32 to 34 months after the event. They found great discrepancies between the two recollections even when participants reported feeling quite confident about their accounts.

In our project, we asked students enrolled in introductory psychology courses to record what they were doing when they first heard of a particular event at two different points in the semester. We collected reports during the first week of the semester and again 2 months later at the onset of the memory section of the course. When we began using this project, we asked about the launching of Operation Desert Storm. For the last several semesters, we have asked about the death of Princess Diana. The students wrote an open-ended description cued by the question “What were you doing when you first heard that the United States had invaded Kuwait [or that Princess Diana had died]?” We encouraged them to be specific and report details if possible.

Following the second memory probe, students analyzed their accounts in class. Brown and Kulik (1977) indicated that flashbulb memories tend to contain particular types of information: (a) location where the information was learned, (b) ongoing activity, (c) informant, (d) affect displayed by others, (e) own affect, and (f) actions taken after hearing the news. Our students examined content of their recollections using Brown and Kulik’s categories. Simply, they determined which of the categories they had recalled. For example, one student reported,

I have a very short memory; I was driving home in my car listening to WMAQ radio after coming out of the Eagle Supermarket and I went into a semi-panic about getting drafted. I bought hamburger patties, buns, fish sticks, curly fries, and Diet Sprite, a good decent meal before eating C-Rations.

We scored this account as including information about location, ongoing activity, informant, and own affect. Next, students compared their accounts for consistency within each of Brown and Kulik’s categories. Most participants (over 70%) reported the same categories of information during both reports. However, the specific details included in each category sometimes underwent change. For example, the previous student described the events somewhat differently during the follow-up.

I still remember every single detail, for I’ve been known to have a razor sharp memory. I remember that I got out of Psychology class and I was going to my Fiancée’s house and it was snowing. I was driving to the Eagle Supermarket where I bought “Mrs. Pauls” 18 pack box of fish sticks and some Ore-Ida Curley Fries. On the way to the store, I turned on WMAQ radio for an up-date; to my surprise, they were sent in to fight.

This student reported the location, ongoing activity, informant, and own affect during both probes. However, the specific details changed. We asked students to examine their own memory reports and determine whether their category accounts completely agreed, agreed with gist, or disagreed. The student described previously could rate his reports of ongoing activity and affect as agreeing with gist whereas the accounts of location and informant completely agreed.

Most probes produced similar accounts (about 30% of category accounts disappeared; about 90% of repeating category accounts completely agreed or agreed with gist). However, some participants produced remarkably inconsistent accounts (about one third of the students produced at least one category that completely disagreed). For example, one stu-
dent reported being home watching television when her mother informed her of Princess Diana's death. In the follow-up probe, the student reported being at her boyfriend's house studying for an exam when he came home with the news. In both of these accounts, the student reported location, ongoing activity, and informant information. Aftermath disappeared in the second probe, and the other three categories of information disagreed.

The project led to lively discussions. Students were pleased and eager to describe their consistent memories as well as their memory discrepancies. One of the most useful aspects of the project is that it provides the class with self-generated examples of many of the concepts covered in an introductory course. For instance, when we want to discuss omission or elaboration, we ask students to search their probes for examples of accounts in which they preserved the gist but added or dropped details. When we discuss schemas, we ask students to search for inconsistent accounts and to speculate on how they may have used an existing schema to reorganize their memories. When examining the effects of emotion, we ask students who had experienced high levels of emotion at the time of encoding to compare their memories with students who had reported little emotion. The same type of discussion can take place when introducing rehearsal. We have found this exercise to be particularly successful for generating examples and illustrating the roles of emotion, rehearsal, consolidation, elaboration, omission, reconstruction, and schemas.

We assessed the effectiveness of this project in two ways: (a) by examining the impact on classroom test scores and (b) by examining course evaluations. We compared scores on examination questions concerning episodic and flashbulb memories (M = 67.84, SD = 22.97) to a sample of factual and conceptual questions concerning other memory phenomena on the same examinations (M = 63.71, SD = 18.09). Students scored higher on the episodic and flashbulb memory questions, t(152) = 2.20, p = .03. In contrast, students enrolled in sections that did not use this project tallied similar scores on these two sets of questions (M = 63.44, SD = 18.34 vs. M = 63.38, SD = 21.97), t(241) = 0.03, p = .97. We also compared scores on the episodic and flashbulb memory questions to performance on those same questions in course sections that did not use this project. Students who had completed this project (M = 67.84, SD = 22.97) scored higher on those questions (M = 63.38, SD = 21.97), t(393) = 1.98, p < .05. Thus, although the effect size is small, it does appear that recording their memories helped the students learn about episodic and flashbulb memories.

We solicited evaluations of the project and the course from students at the end of the semester. Students rated the course as 4.4 and the project as 4.6 on a 5-point scale ranging from 1 (poor) to 5 (outstanding). Written comments were also uniformly positive. Students described the project as "intriguing" and "interesting." They were especially pleased to be "able to analyze 'real data.'" Students also reported that they believed the project helped them to better understand memory processes.

Overall, we believe this project is rewarding and instructive for both faculty and students. The project improved student involvement and learning. It required just a few minutes of class time to collect the memory probes and as much time in discussion as the instructor desires. It is a project that can be performed every semester. Many events occur that might trigger vivid memories. For example, recent flashbulb memory research (for a review, see Conway, 1995) has examined memories for the assassination attempt on President Reagan, the San Francisco earthquake, and the Los Angeles earthquake, among others. Although events such as these do not occur at the onset of every semester, we have been able to use the same event for several semesters until another unexpected and emotional event occurs. Finally, we believe that this project is flexible enough to be useful in a number of courses. In introductory psychology we have used it as a vehicle for discussion of memory phenomena. We have also used this project in upper level learning and memory courses. Here, we require more extensive data analysis, a literature review, and completion of a laboratory paper. As a result we have been able to impart concepts in research design and analysis as well as teach about memory phenomena.

References


Note

Send correspondence to Mark Sudlow Hoyert, Psychology Department, Indiana University Northwest, Gary, IN 46408; e-mail: mhoyert@iunhaw1.iun.indiana.edu.

Demonstrating the Concept of Illusory Correlation

Jay W. Jackson
Indiana University–Purdue University

The concept of illusory correlation helps explain how some social stereotypes are partly due to basic cognitive processes. I describe a classroom activity that essentially replicates, in a simplified format, a classic study of the illusory correlation effect. In addition to effectively demonstrating the phenomenon of illusory correlation, this activity provides a stimulus for discussing related social psychological issues and methodological procedures.

Research has demonstrated that people pay special attention to distinctive or rare events and recall such events with
negative attitudes toward aviation had an equivalent impact on annoyance regardless of noise exposure. These findings strongly suggest that the association between attitudes and annoyance is partially independent of noise exposure (although the causal direction between them remains unclear).

Suggestions for Avoiding Confusion

The intended meaning of the phrase “correlation does not imply causation” is important, and the confusion that is apt to result from it may be detrimental by contributing to inappropriate conclusions from research results. Thus, all teachers of psychology should take care to avoid the confusion surrounding the phrase by adopting the following recommendations:

1. Employ the term correlation only in a statistical context, and then refer specifically to correlational analysis or the correlation coefficient.
2. Employ the term association to describe the relation between two variables without identifying the particular statistic employed to detect the relation.
3. Refer to research designs in which researchers merely observed variables (rather than manipulating them) as nonexperimental designs rather than correlational designs.
4. Replace phrases like “correlation does not imply causation” with the phrase “without manipulation, association does not imply causation.” Given people’s predilection for catchy phrases, it appears wise to offer a suitable alternative to the flawed but commonly used existing phrase. Our alternative phrase correctly stresses the importance of the type of research design, rather than the type of statistical analysis performed. Nonetheless, a more substantial phrase such as “association without manipulation does not normally imply causation” would better account for circumstances in which causal inferences from nonexperimental data are defensible. A blander phrase, such as “nonexperimental designs do not imply causation” suffers from not being sufficiently catchy.

Conclusions

In our experience, many students mistakenly form the impression during their undergraduate years that limitations on causal inference are imposed by the correlation statistic rather than by nonexperimental research designs. Although objective data are not yet available, casual observation suggests that through application of the recommendations offered in this article, teachers of psychological statistics and research design can help to eliminate such errors and facilitate more appropriate inferences from nonexperimental and experimental research.

References


Notes

1. R. F. S. Job is currently General Manager of Road Safety Strategy Branch, Road and Traffic Authority of New South Wales, Australia.
2. Send correspondence to Julie Hatfield, NSW Injury Risk Management Research Centre, University of New South Wales, New South Wales, 2052, Australia; e-mail: J.Hatfield@unsw.edu.au.

An Active Learning Classroom Activity for the “Cocktail Party Phenomenon”

Michael A. Clump
Marymount University

This article presents an active learning demonstration of the “cocktail party phenomenon.” It involves dividing the students in the class into groups of 3, with 2 individuals acting as speakers and 1 person as the participant. By simultaneously involving all of the students, more students experience the effect in an environment that replicates the cocktail party phenomenon and the students’ experiences outside of class, such as in a coffee house. The 3 within-subjects conditions illustrate how certain information slips through an attentional block. The students’ evaluations indicated the technique was enjoyable, useful, and a good way to learn about the topic.

Everyone experiences a time when they “tune out” someone else while talking on the phone or watching television. In addition, everyone listens at a party to one person tell a story, but occasionally eavesdrops on another conversation. Cherry (1953) and Cherry and Taylor (1954) began the study of selective attention by using a laboratory task in which participants immediately repeated information presented to one ear, known as shadowing, and ignored information presented...
to the other ear. The task, known as the dichotic listening task, involved presenting auditory information to participants through a pair of earphones. As a result, researchers could alter the information to be shadowed by the shadowed ear and the information to be ignored by the nonshadowed ear. Cherry found that participants recognized when information presented to the nonshadowed ear was speech or when a male or female voice presented the information. Moray (1959) found that participants remembered little of the information presented to the nonshadowed ear, unless their names prefaced the information, such as instructions. Consequently, participants in a dichotic listening task remember little of the nonshadowed information because it receives less attention.

A participant’s experience in the laboratory dichotic listening task compares to an individual's experience at a cocktail party. At a cocktail party, a person will be actively involved in a conversation while other conversations occur. The individual remembers the information from the engaged conversation, but cannot remember information from the other conversations. The caveat to this scenario occurs when the person hears his or her name, or other information related to the individual, spoken in one of these surrounding conversations. Like the participants in Moray’s (1959) study, the individual diverts attention to the surrounding conversation. A cocktail party is a real-world example of the laboratory findings from the dichotic listening task, and thus, a second name given to the dichotic listening task is the cocktail party phenomenon.

Goodwin (1988) described a dichotic listening classroom activity that used students to present auditory information instead of laboratory earphones. Goodwin’s activity involved two students who presented the shadowed and nonshadowed information to 4 or 5 separate participants, with each participant replicating one of Cherry’s (1953) or Moray’s (1959) conditions. For instance, 1 participant heard one of the presenters alter his or her pitch, another participant heard a distracting word repeated periodically, and a third condition included the participant’s name in the nonshadowed passage.

I modified the laboratory dichotic listening task and Goodwin’s (1988) classroom demonstration of this laboratory task to create an easy-to-replicate in-class demonstration of the cocktail party phenomenon. I developed this classroom activity to help students connect what they experience outside of class, such as at parties, at coffee houses, and at other social events, where they are in a conversation and overhear other conversations, with the laboratory findings discussed in class.

This demonstration involves all of the students working simultaneously, so all of the students actively learn about the topic through their experiences, instead of only a few students as in Goodwin’s (1988) demonstration. Although having all of the students concurrently involved in the demonstration may appear chaotic, these multiple distracting inputs help them learn through an active, direct replication of their real-world experiences in their dorm rooms, in the cafeteria, in restaurants, or in coffee shops, which may not occur when involved in a replication of the laboratory task, such as in Goodwin’s demonstration. Consequently, this activity fits with the design principles Mathie et al. (1993) described for an active learning activity.

My technique allows for the information presented to be altered for each participant, by incorporating first names, middle names, last names, nicknames, friends’ names, common classes, and common experiences, instead of using a prerecorded, nonpersonally relevant, message as a direct replication of the laboratory task, such as Goodwin’s (1988) technique would use. Thus, as students actively develop and execute this demonstration, they connect the demonstration and the discussion about the findings to their specific lives via self-reference, which Schmeck and Meier (1984) found to be a highly effective way for students to learn and remember material, and the creation of a meaningful context as Mathie et al. (1993) described. This active development of the material, which is different from using predeveloped, general stimuli from a direct replication of the laboratory task and Goodwin’s demonstration, also provides students with an opportunity to actively design and conduct research in cognitive psychology.

**Description of the Activity**

The activity involves dividing the entire class of students into groups of 3 students independent of gender, such that my typical class with 30 students would work in 10 groups. The activity uses a within-subjects design with three conditions. Each group of students decides who will act as the Participant the two Speakers. The Participant in the activity always shadows Speaker 1 and tries to ignore Speaker 2; Speaker 1 tells an interesting, but different story to the Participant during the three conditions; Speaker 2 also says information to the Participant, but what Speaker 2 says is different in the three conditions. The groups of students arrange themselves so that the Participant sits between the two speakers. The Participants leave the room for a few minutes as I describe the demonstration to the students who will serve as the two speakers for their groups.

Although all of the groups simultaneously conduct the activity, for ease of presentation I describe the activity of one group. Once the Participant exits the room, I assign one student to be Speaker 1 and the other student to be Speaker 2. After explaining Speaker 1’s role, which involves continuously telling a story the Participant repeats during each condition, I describe the role of Speaker 2 (i.e., the nonshadowed speaker), who alters the dictated information, based on the following within-subjects conditions:

**Condition 1:** Speaker 2 says random words and numbers at 1 word/sec, without including any interesting information.

**Condition 2:** Speaker 2 again says random words and numbers at 1 word/sec, occasionally including the name of the Participant.

**Condition 3:** Speaker 2 tells an interesting story that includes the Participant’s name and information from school, classes, friends, jobs, relatives, or other information specifically relevant to the Participant. (Because students work together on other...
group projects and assignments in the course, this type of information typically becomes easily available.

The two Speakers develop the stories and word lists they will use during the three conditions as the Participant enters class. The Participant sits between the other two group members, both of whom face the Participant and talk directly into the Participant’s ears from less than 1 ft away once the demonstration starts. I then explain how the Participant will immediately repeat exactly what Speaker 1 dictates, while ignoring what Speaker 2 says. Finally, I explain that the Participants will be involved in the task three separate times for 1 min each.

After 1 min of the activity, the group stops and the Participant writes down for 1 min exactly what he or she recalls from the shadowed and nonshadowed ears, during which Speaker 1 and Speaker 2 refresh their stories and word lists. The group repeats the procedure for all three conditions.

Following the demonstration, I ask all Participants to describe the information that they recalled from both the shadowed and nonshadowed ears for the three conditions. Invariably, the Participants describe how their recall was excellent for the shadowed material for Conditions 1 and 2, but the shadowed information recalled for Condition 3 was substantially reduced. In contrast, the Participants’ recall of the nonshadowed information significantly increased from Condition 1 (e.g., very little), to Condition 2 (e.g., their names multiple times), to Condition 3 (e.g., most of the material presented by Speaker 2). For each Participant, the entire activity required between 15 and 20 min; consequently, we repeat the activity so the students can rotate through the three roles during a class period and experience the activity from each role.

Evaluation of the Activity

Students (N = 22) evaluated the activity on six questions adapted from Madson (2001). Over 90% of the students found the demonstration to be either enjoyable or very enjoyable, and all of the students agreed or strongly agreed with using the demonstration in the future. All the students found the demonstration to be useful or very useful for prompting them to think about the cocktail party phenomenon and dichotic listening. Nearly 91% of the students believed they learned more about the topic having participated in the demonstration. The students did not feel that the demonstration was a waste of their time, as evidenced by the fact 95.5% selected strongly disagree or disagree. Finally, over 80% of the students selected agree or strongly agree for the statement that other instructors should use the demonstration when teaching about the topic in other related courses.

Discussion

The comments made by the students indicated that they enjoyed the activity, especially with the personally relevant information. The Participants indicated they wanted to keep listening to the nonshadowed information, especially when it contained gossip, information related to a course, or a funny story. The students stated that the activity helped them understand Cherry’s (1953), Cherry and Taylor’s (1954), and Moray’s (1959) laboratory findings on selective attention. An instructor could also use the activity to discuss how to design and conduct research in cognitive psychology. The students could redesign the demonstration to effectively measure their inclinations to listen to the nonshadowed information, especially the personally relevant information, which would further involve the students in actively learning about these research studies.

The activity described by Goodwin (1988) and my activity provide faculty with separate demonstrations of selective attention. Whereas Goodwin’s technique provides an excellent method to specifically replicate the laboratory nature of the dichotic listening task, my technique focuses on tying dichotic listening to the students’ real-world experiences, like the cocktail party phenomenon, by attempting to connect what students experience outside of class with research studies discussed in class. In addition, all of the students benefit from being involved in the active learning focus of the demonstration. The focuses of the activity, involving all of the students in the demonstration and connecting it with their everyday life by having them apply, evaluate, and synthesize the topic, match the main principles of active learning activities espoused by Mathie et al. (1993).

References


Notes

1. I thank Randolph Smith and the three anonymous reviewers for their helpful comments on earlier versions of this article.

2. Send correspondence to Michael A. Clump, Department of Psychology, Marymount University, 2807 North Glebe Road, Arlington, VA 22207; e-mail: michael.clump@marymount.edu.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
An Effective Exercise for Teaching Cognitive Heuristics

Alan Swinkels
St. Edward’s University

This article describes a brief heuristics demonstration and offers suggestions for personalizing examples of heuristics by making them relevant to students. Students complete a handout asking for 4 judgments illustrative of such heuristics. The decisions are cast in the context of students’ daily lives at their particular university. After the professor tallies responses to each question, students discuss their choices and the reasons for them during a subsequent class meeting. Student feedback indicated that this exercise is informative, interesting, and enjoyable. Results from a pretest–posttest study confirmed that student learning improved from before to after the administration of the exercise.

Social perceivers rarely have the luxury of gathering and weighing all the pertinent evidence needed to make a social judgment. Rather, they typically must balance the twin demands of efficiency and accuracy when judging others (Fiske & Taylor, 1991). Because the social world is a busy place with many elements competing for attention, people must be able to make quick and accurate social judgments. In short, we need to be right, but we also need to be fast when sizing up our social world.

There are several theoretical viewpoints addressing how people perform acts of social perception (for an overview, see Fiske & Taylor, 1991; Higgins & Bargh, 1987; Kelley, 1967; Showers & Cantor, 1985). Among these views, the cognitive miser approach (Taylor, 1981) argues that because people are limited in their capacity to process information, they adopt shortcuts whenever possible. To that end, social perceivers often rely on cognitive heuristics when judging others. Although heuristics help us achieve a balance between efficient social perception and accurate social perception, they can lead us astray. An overreliance on representativeness, availability, simulation, or the anchoring and adjustment heuristic, for example, can lead to conclusions that are incorrect when compared with a more careful, reasoned analysis of the evidence (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1973, 1974).

Much of the empirical work on heuristics involves presenting research participants with short scenarios calling for social judgments (e.g., asking people to judge the political affiliation of a described person). Although these scenarios clearly capture heuristic reasoning at work, they can be modified to become more personal and memorable for students in a classroom setting. The exercise I describe presents a short, effective demonstration of several heuristics.

The Exercise

Administering the Exercise

In my social psychology course, I discuss heuristic reasoning during the first 2 weeks of class in the context of person perception, impression formation, and social cognition. At the end of the class meeting on the day before we discuss heuristics, I distribute a handout (see Appendix), telling students that “it’s a simple exercise in making judgments about others that will help us discuss our next topic.” In the context of the week’s material the request seems natural, and students are not aware that the judgments they make will invoke the use of heuristics. The handout takes less than 5 min to complete. I ask students to write their names on their sheets (so I can return them later) and I collect their responses before class adjourns.

Distributing the handout prior to discussing the topic produces several benefits. First, I can tally students’ responses and adapt my presentation based on patterns that emerge. Second, if students catch on to these heuristics at work and try to change their answers during class discussion, I can good-naturedly produce the prior evidence of their written responses. Finally, students are eager to return to class the following day to discover what the exercise illustrates.

Elements of the Exercise

The handout presents four judgments that illustrate the use of the representativeness, availability, simulation, and anchoring and adjustment heuristics. In each case, relying on a cognitive shortcut produces a less-than-optimal decision.

Representativeness. In making this judgment, students rely on the fact that a description of a specific individual seems to be a good representation of a larger social category (e.g., Rudy’s peculiar habits are suggestive of a performer). However, the base rate (i.e., the statistical probability) of any of the other options (e.g., lawyer, surgeon) is substantially higher, making it more likely that the person described holds one of those other occupations. Social perceivers typically ignore the base rate information and are swayed by the seeming representativeness of the description instead (e.g., Kahneman & Tversky, 1973).
**Availability.** This example asks students to judge whether more words in a novel are likely to end in *-ing* or to have *n* as the second-to-last letter. The availability heuristic leads students to generate examples of each category to help make their decision. Because examples from one category are more available in memory (running, jumping, flying, skiing, driving, hiking) than are examples from the other (sink, examine, blond), the first case is judged more likely. However, all *-ing* endings contain *n* as the second-to-last letter, making it easy to demonstrate that *n* endings typically are more likely to occur, despite students’ intuitions.

**Simulation.** This example asks students to judge which of two college roommates would be more upset by being denied permission to enroll in a needed course. Logically, neither student has more cause to be upset. Each student is shut out, both students have to start over, and neither student has much recourse. However, people overwhelmingly judge the student who missed the opportunity by 10 min to be more upset than the student who missed the opportunity by an entire day. The simulation heuristic relies on the ease of constructing hypothetical scenarios or alternative outcomes. Specifically, because it is easier to imagine ways that a 10-min gap could have been changed than a full day’s gap, people give the first occurrence more weight in influencing their judgments about emotional reactions to these events. This question is a variation of Kahneman and Tversky’s (1982) Mr. Crane and Mr. Tees example. It is more relevant to students because the outcomes are in terms of a college experience that many students have likely encountered.

**Anchoring and adjustment.** The final example asks students to make a judgment under ambiguous conditions. When faced with this kind of task, most people look for an “anchor” or reference point on which to make their decision and then adjust their final judgment accordingly. However, adjustments are usually insufficient (remaining close to the initial anchor point), and the anchors themselves may not be relevant to the judgment at hand.

For this demonstration, I prepare two versions of the handout. Both are identical, except that half the handouts give a low anchor for this question and the remainder give a high anchor. For example, when asking students to judge how many academic departments are at their university (an ambiguous judgment for which most students have no point of reference) I supply a reasonable low anchor (e.g., 15) on half of the handouts and a reasonable high anchor (e.g., 30) on the other half. Students typically adjust their judgments around the anchor provided on the handout.

**Discussing the Exercise**

After students have completed the exercise, I devote the subsequent class meeting to considering heuristics and their pitfalls. After a brief introduction, I return the handouts to the students, and we discuss them for the remainder of the class.

**Emergent Patterns**

I have used this exercise successfully in my social psychology, introductory psychology, and cognitive psychology courses during the past 14 years. Almost without exception, the questions lead students to rely on the intended heuristic when making their judgments. For example, students think peculiar cousin Rudy is most likely to be a trapeze artist, although a few savvy students occasionally choose occupations with a higher base rate (e.g., lawyer). Similarly, students generally think Mario is more upset by barely missing his class than Victor, illustrating use of the simulation heuristic. Interestingly, there is often an equal division of students judging the endings of words in a novel. Discrepancies favor the operation of the availability heuristic (i.e., selecting the *-ing* ending as more common), but perhaps because this judgment is less “social” and presented with a big hint (*-ing versus *-n-*) , the disparity between judgments is not as great as it is in other cases. This result is in contrast to the anchoring and adjustment example, which clearly shows the operation of that heuristic. When I compute the means of the estimates for each version, it is clear that students rely on the arbitrary anchor provided and that their adjustments remain close to it (e.g., 15.2 and 27.9 when given anchors of 15 and 30).

**Student Comments**

When discussing the handout, students often provide “textbook” definitions of the heuristics even as they defend their misbegotten choices. When asked, for example, why Rudy is so likely to be a trapeze artist, at least one student will respond, “Well, he sounds like he’s a trapeze artist,” which is the epitome of relying on representativeness. The handful of students who relied on base rate information usually make their case quite persuasively, and all students soon understand the operation of the heuristic. In short, students engage in lively discussion over their choices and come to realize that heuristic shortcuts can sometimes falter in comparison to a more reasoned analysis.

**Evaluating the Effectiveness of the Exercise**

Using scales ranging from 1 (not at all) to 7 (very much), students have rated various aspects of the social psychology course at the end of each semester over the past several years. Students invariably rate this exercise as being quite effective. Aggregating the mean ratings for each question across seven sections of the social psychology course (total N = 198), students agreed that (a) the presentation helped them understand related text and lecture material (M = 5.84); (b) it was informative (M = 5.87), interesting (M = 5.72), and enjoyable (M = 5.62); and (c) the exercise should be retained in future sections of the course (M = 5.87). In every case the aggregated average rating is significantly different from 4, the neutral midpoint of the rating scale (all one-sample t tests significant at p < .01, one-tailed).
Thirty-five students (24 women, 11 men) enrolled in introductory psychology (n = 22) and social psychology (n = 13) courses during the Spring 2001 semester completed an identical pretest and posttest to assess the effectiveness of the exercise. I administered the pretest a week before the exercise and the posttest a week after the exercise. Both measures presented eight definitions of social behaviors and students selected which of 12 available terms described each one. I embedded the four target heuristics (representativeness, availability heuristic, simulation heuristic, anchoring and adjustment) among four filler items (self-verification, conjunction fallacy, false consensus effect, primacy effect) in the set of definitions. The list of available terms also included fundamental attribution error, belief perseverance, self-fulfilling prophecy, and hindsight bias.

A paired-samples t test comparing the mean number of correct identifications of the four heuristics items on the pretest (M = 1.77, SD = 1.19) and the posttest (M = 3.26, SD = .82) indicated a significant improvement in students’ ability to identify these concepts, t(34) = 6.37, p = .0001 (two-tailed), η² = .54. An examination of the modes showed that students correctly identified two of the heuristics (50%) on the pretest and four (100%) of the heuristics on the posttest. These data suggest that accuracy improved from before to after the presentation of the exercise. However, students were no more accurate on the four filler items on the posttest (M = 1.74, SD = 1.12, mode = 2) than they were on the pretest (M = 1.80, SD = .93, mode = 2), t(34) = .31, ns. Students in the social psychology course showed a small (M = 3.62, SD = .65) but significant advantage over students in the introductory psychology course (M = 3.05, SD = .84) in the mean number of correct identifications on the posttest, t(33) = 2.09, p = .04 (two-tailed), η² = .12.

Conclusions

This brief exercise is easy to administer, engaging for students, and effective in demonstrating heuristic reasoning. Moreover, it can be used successfully in a variety of courses, such as social psychology, introductory psychology, and cognitive psychology, or any course that considers human reasoning and its limits. Creatively tailoring the details of the exercise to incorporate the experiences of students at their university can make it more engaging.

References


Appendix

An Exercise in Social Judgment

1. Dr. Swinkels’ cousin, Rudy, is a bit on the peculiar side. He has unusual tastes in movies and art; he is married to a performer, and he has tattoos on various parts of his body. In his spare time Rudy takes yoga classes and likes to collect 78 rpm records. An outgoing and rather boisterous person, he has been known to act on a dare on more than one occasion. What do you think Rudy’s occupation most likely is?
   A) Farmer  B) Librarian  C) Trapeze Artist  D) Surgeon  E) Lawyer

2. In one chapter of a best-selling novel, would you expect to find more words that (circle one)
   a) end in -ing (—ing) or b) have n as the second to last letter (——n—)?

3. Two college roommates, Victor and Mario, are registering for courses for the spring semester. They leave their dorm room together, stop and eat breakfast together, chat with a mutual friend, and arrive at the registrar’s office at the same time. They both line up to enroll in their art history classes. Victor is told the class he wanted was filled to capacity at the end of the previous day. Mario is told the class he wanted was filled to capacity 10 minutes before he arrived. Who is more upset, Victor or Mario?

4. How many academic departments are there at St. Edward’s University? (circle one)
   a) Fewer than 30  b) More than 30
   What is your exact guess? Write a number on this blank line: ______________

Notes

1. Some of these results were presented at the 21st annual National Institute on the Teaching of Psychology, St. Petersburg Beach, FL, January 1999.
2. I thank Traci A. Giuliano for her helpful comments on earlier drafts of this article.
3. Send correspondence to Alan Swinkels, Department of Psychology, St. Edward’s University, 3001 South Congress Avenue, Austin, TX, 78704; e-mail: alans@admin.stedwards.edu.
Central to any discussion of human reasoning and decision making is the idea that humans do not process information objectively. Humans are cognitive misers who typically base their decisions on mental shortcuts and heuristics instead of on the laws of logic, statistics, or probability (e.g., Kahneman & Tversky, 1973, 1982, 2000). Although heuristics are generally quick, efficient, and accurate; they occasionally produce systematic errors in people’s reasoning and decision-making processes. Generally, textbook authors (e.g., Nairne, 2003; Weiten, 2003) choose to illustrate this concept using standard heuristics such as availability and representativeness (Kahneman & Tversky, 1973, 1982, 2000). Although these are excellent demonstrations, I have found that a demonstration of the Monty Hall Dilemma (MHD) clearly illustrates the aforementioned principles and can be used effectively in any course that includes coverage of thought processes (e.g., introductory psychology, cognitive, social), research methodology, or probability (e.g., research methods, statistics).

The Monty Hall Dilemma is based on the classic game show Let’s Make a Deal (1975–1985), which starred Monty Hall as the amiable host. On the show, Hall invited a contestant to participate in a game of chance in the hope of winning a magnificent prize. The contestant chose one of three doors (1, 2, or 3) and knew that there was a prize behind only one of the doors. The contestant tried to select the winning door (e.g., choose 3). Monty Hall, always the showman, then made the game more interesting by revealing that no prize resided behind one of the doors (e.g., 2). At this point, only two doors remained (1 and 3), and the contestant made a crucial choice—to stay with the original selection (3) or to switch to the remaining door (1).

When the game began, most contestants realized that they had a 1 in 3 chance to pick the correct door (33.3%). When Monty offered the final choice, however, many contestants believed that the odds had changed in their favor. Given that only two doors remained, they assumed there was now a 1 in 2 chance to pick the correct door (50.0%). Thus, when asked to stay or switch, the contestants argued that the choice was irrelevant—after all, they appeared to have a 50–50 chance either way. Unfortunately for the contestants, their reasoning was flawed—the switch strategy actually leads to more wins than the stay strategy (e.g., Granberg & Brown, 1995; Mosteller, 1965; Selvin, 1975).

Most contestants, and indeed people from all over the world (Granberg, 1999), have difficulty accepting that either strategy should have an advantage in this situation. One, however, can clearly demonstrate that the switch strategy is more profitable. At the beginning of the game, the contestant picked a door randomly (e.g., 3) and had a 33.3% chance of being correct. Consequently, there was a 67.6% chance that the prize was not behind the selected door (i.e., it is behind one of the remaining doors—1 or 2). If the contestant did nothing or stayed, he or she would be correct 33.3% of the time. When Monty eliminated a door (2), the original prob-

---

### Demonstrating the Monty Hall Dilemma

Matthew R. Kelley  
Lake Forest College

This article describes 2 simple and effective classroom demonstrations of the Monty Hall Dilemma (MHD). The MHD is based on a choice scenario from the game show Let’s Make A Deal (1975–1985) in which a contestant attempts to select the winning door from a set of 3 alternatives. The dilemma emerges when, after the initial selection, 1 of the nonwinning doors is revealed and the contestant must choose whether to stay with the initial selection or switch to the remaining door. Although intuitively it seems that the odds of winning are now 50–50, in reality, switching produces more wins than staying. The MHD illustrates the pitfalls associated with the use of cognitive shortcuts or heuristics and can be used to demonstrate the importance of empirical observation and experimentation.
lem state did not change. There was still only a 33% chance that original choice was correct; however, now there was a 67.6% chance the remaining door (1) was correct. This solution has been proven mathematically (e.g., Selvin, 1975), using Bayes Theorem (e.g., Diaconis & Zabell, 1986), and through computer simulation (Shaughnessy & Dick, 1991). It can also be demonstrated experimentally in a class using a sample of students as contestants.

Activity Materials and Procedure

Typically, I employ two different versions of this demonstration—one for my research methods and statistics laboratory and one for my cognitive class. However, one can easily adapt this activity for almost any class, regardless of its size.

In my research methods and statistics laboratory, I use the MHD as I introduce the importance of gaining knowledge via empirical observation and experimentation as compared to relying on simple reasoning or intuition. After introducing the game and having a student participate as a contestant, I ask the class to predict whether one strategy would be more effective than another. Without fail, the class responds that the choice is moot because the odds are 50–50 with either strategy. Following a thorough explanation of the true nature of the MHD, I typically find that most of the class continues to trust their intuition and reasoning over my explanation. At this point, I suggest that the class resolve the issue by conducting an experiment in which they systematically compare the two strategies by simulating the game show and recording their own data.

In advance, I prepare a general data collection sheet that consists of six columns (trial number, winning door, contestant’s first choice, door opened by host, contestant’s final choice, and contestant’s outcome). Each student receives two data sheets (one for each strategy) and completes 84 game simulations (42 trials for each strategy). Although I prefer to use a larger number of trials in my laboratory, one can use as few as 20 trials for each strategy to demonstrate the MHD. Prior to the start of the experiment, each student prepares the information required for his or her turn as host by selecting a winning door for each trial using a random numbers table. Once the doors are determined, one student serves as host (solicits initial door choice from contestant; reveals nonwinning door; solicits final door choice; reveals outcome) and records all the relevant data for each trial; the other student serves as the contestant (randomly selects first door, adopts strategy, makes final choice). At the completion of the 84 trials, the students switch roles and repeat the procedure. Finally, the students calculate their individual percentages of wins produced by each strategy to determine whether the switch strategy was indeed more effective than the stay strategy; these percentages are then pooled for the class and used in the discussion and lab report. On average, switching should produce wins 67% of the time and staying should produce wins only 33% of the time.

In my cognitive class, I use the MHD, along with other common heuristics and fallacies (e.g., availability, anchoring, conjunction fallacy, base rate neglect), to demonstrate that humans are not rational decision makers. After gathering predictions and explaining the true nature of the MHD, I provide a short demonstration of the effect by randomly selecting 20 students to serve as contestants and 1 student to serve as my cohost (to record the data and to ensure that I run an honest game). I quickly administer two rounds of the game for each student—one in which the student is instructed to stay and one to switch. In a group of 20 students, one should expect between 12 to 14 wins for the switch strategy and between 6 to 8 wins for the stay strategy.

In a recent research methods and statistics course, I surveyed my students about their perceptions of the Monty Hall Dilemma demonstration using a survey modeled after Munro and Munro (2000). They rated the activity on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree), for each of four statements (see Table 1). Overall, the students considered the MHD to be highly interesting and useful and clearly recommended the use of the activity in future classes. The MHD demonstration is easily adaptable to the needs of any class, requiring as little as 10 min of class time and as much as 50 min (lab activity). In place of an in-class demonstration, one could assign the task as a homework assignment or point students toward a Web-based simulation of the MHD (e.g., CogLab; Francis, Neath, & Surprenant, 2000). Regardless of the specific procedure used to demonstrate the MHD, it will serve as an interesting, enjoyable, and effective demonstration for any discussion of human thought processes, research methodology, or probability.

References


Table 1. Student Evaluation for Monty Hall Dilemma Demonstration

<table>
<thead>
<tr>
<th>Statement</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. This exercise was an interesting learning experience</td>
<td>4.28</td>
<td>0.57</td>
</tr>
<tr>
<td>2. Doing this exercise was a valuable way to learn</td>
<td>4.28</td>
<td>0.67</td>
</tr>
<tr>
<td>3. This exercise helped me to understand the importance of systematic observation in gaining knowledge</td>
<td>4.28</td>
<td>0.57</td>
</tr>
<tr>
<td>4. I would recommend this exercise for future classes.</td>
<td>4.50</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Note. Each statement was rated on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree).
Themes and Principles of Child Development Illustrated in Music

Marvin W. Daehler and Kristen E. Miller

University of Massachusetts at Amherst

To encourage students to experience key themes and principles of child development in a popular medium, we played musical selections ranging from contemporary rock to children’s pieces before each session of a child development class. Simultaneously, visual projection of the lyrics along with text and pictures highlighted the relevance of the music to that day’s discussion. Students indicated the audio–visual presentations were interesting and pertinent and should continue in future offerings of the course. The activity provides students with opportunities to become aware of themes and principles of development from a socially and culturally significant resource not routinely considered by instructors.

Instructors teaching large sections of courses on child development and related topics may seek to identify opportunities for students to become acquainted with themes and principles of behavior from contemporary media. Anecdotes recount the experiences and functioning of children as seen through the eyes of literary experts (e.g., Landau, Epstein, & Stone, 1972; Sattler, Kramer, Shabatay, & Bernstein, 2000), and commentaries discuss how film and television portray children (e.g., Tylim, 1997; Wolfenstein, 1955). Documenting the benefits of both live-action and animated feature films, Boyatzis (1994) and Kirsh (1998) reported positive student interest and pedagogical value in promoting greater understanding of social and personality development from these media. Green (2003) highlighted effective ways to use film and video both within and as a supplement to the classroom. However, no published information addresses the use of contemporary music as a resource to instructors of child development courses.

Although they may have preferences, many students are familiar with a wide variety of styles of music. We identified and used assorted musical pieces to highlight concepts under discussion in the child development class. The purposes were not only to reveal to students themes and principles relevant to child development in music but also to prepare them for the course content for that day. We played a musical selection immediately prior to the start of each class period. Accompanying each selection were PowerPoint™ slides showing the lyrics of the song along with images and other text accentuating ideas relating to the day’s topic. We report students’ interest in and perceived value of this effort.

Musical Selections and Presentation

We first identified 13 primary topics typically included in child development textbooks and likely to be covered in class. Library resources and Internet searches of lyrics of songs (using, e.g., key words such as music, child, and baby and a relevant psychological term), as well as our personal knowledge and that of our colleagues, usually led to identifying several pieces of music related to each topic. Selections represented a wide range of songs performed by popular artists, classic and contemporary rock stars, folk singers, and country musicians, pieces produced in Broadway musicals, and a few songs specifically designed for children selected from Sesame Street© or from the creative works of other performers. Table 1 lists topics, a relevant song and artist for each, and a brief indication of how each of the pieces relates to the course content.

We played one selection immediately prior to the start of each class. The length of the selection dictated when the music and slides began, but each piece ended by the scheduled beginning of class. Students were free to listen and watch the slides or to engage in other activity including socializing with classmates or communicating with the instructor as the music played. Thus, we designed the activity to supplement rather than be a required component of the course.

Anywhere from 5 to 12 PowerPoint® slides accompanied the various selections depending on lyrics and the duration of the music. On average, slides were projected for about 20 sec each, but typically ranged from 10 to 40 sec. To illustrate the kinds of information shown on the slides, the material accompanying the playing of “Luka,” a song about a physically abused child, included definitions of and statistics relating to the frequency of child abuse in the United States as well as factors correlated with the occurrence of child abuse in families. The slides accompanying “Do-Re-Mi” described developmental differences in the use of mnemonic techniques with emphasis on elaboration, the technique evident in this piece. The slides shown with “Itsy Bitsy Spider/Coming Around Again” outlined the distinction between perceived helplessness and mastery orientation and highlighted the necessary role of effort for achievement and success. In the case of “Mairzy Doats,” the slides included spectrograms of an individual saying the text of the song to illustrate the cues associated with, and the difficulty of identifying, segmental phonemes in the stream of speech.
taken my course and forgotten the phrase “regression line,” they have described the scatterplot exercise to me, how the students were clustered along the diagonal of the room, and the relation between height and shoe size that was reflected in the fact that there were no short students with very big feet or tall students with very small feet.

References


Note

Send correspondence to Jane Marantz Connor, Department of Psychology, Binghamton University, Binghamton, NY 13902–6000; e-mail: jconnor@binghamton.edu.

Helping Students Gain Insight Into Mental Set

Richard A. Griggs

University of Florida

Difficulty in problem solving is an important topic in the thought-language chapter in introductory textbooks. Textbook authors typically use water-jar problems to illustrate mental set, one of the major barriers to successful problem solving. Such problems, however, only exemplify mental set created within the actual problem setting. To extend such discussions, I describe an engaging class activity that uses 4 related series problems to illustrate mental set and its more general negative effect on problem solving. Student reports indicated that they found the activity both engaging and instructive. I also provide resource references to extend class discussion of the activity to the broader issue of mindfulness versus mindful behavior.

Difficulties in problem solving comprise a key section in the thinking-language chapter in introductory textbooks. Two barriers to problem solving are functional fixedness (viewing the objects in the problem environment as having fixed functions) and mental set (repeating the same solution strategies used in the past for similar problems). Textbook authors typically illustrate these two concepts by using two mental-puzzle type problems studied by Gestalt psychologists, usually Duncker’s (1945) candle mounting problem and Luchins’s (1942) water-jar problems, respectively.

Although Matlin (1999; see also 2002, pp. 380–383) employed these two Gestalt problems as examples of functional fixedness and mental set, she discussed these difficulties as examples of a more encompassing hindrance to problem solving—Langer’s (1989) concept of mindlessness. As Matlin pointed out, mindlessness is a form of automatic thinking in which people use information too rigidly, look at a problem from only one perspective, and rely too heavily on old information at a cost of not being aware of new information in the environment.

In addition to a set of water-jar problems, Matlin (1999) used a series problem to illustrate mental set. Series problems require the solver to induce the structure of a series of numbers, words, or symbols and then project the next item in the series. The structure of Matlin’s series problem required a novel solution approach. Thus, repeating the typical strategies for past series problems (mental set) hinders the solution process. The interfering past experience relevant to water-jar problems stems from solving previous water-jar problems in the current set of problems. For series problems with novel structures, the interference arises from past solution experiences with these problems outside of the current setting. In addition, whereas mental set usually prevents only the most efficient solutions to water-jar problems, it typically prevents solution for these novel series problems.

The activity described in this article expands Matlin’s use of a novel series problem to illustrate the broader sense of mental set by including four such problems that all have essentially the same, very familiar solution but remain independently difficult because each requires a different novel solution strategy. The nature of the interfering mental set therefore varies across the four problems. Thus, these problems both exemplify the general nature of mental set and its key role in making problems more difficult and set the stage for a discussion of the more general concept of mindlessness.

Activity Materials and Procedure

For a better understanding of this activity, you should attempt to solve the four series problems given in Table 1 before reading further. One can present the four problems on a chalkboard, overhead transparency, or Microsoft PowerPoint slide. Present each problem one at a time and in the order given in the table. As you proceed from one problem to the next, leave all prior problems in view and do not give any of the answers until you have presented all of the problems.

Ask students who have seen a problem before or solve a problem not to give the answer, but rather to think of clues to help their classmates overcome mental set and gain insight into the problem’s solution. These students can also testify that the problems all have rather simple answers, which keeps other students motivated to search for solutions. In addition, the students providing clues gain a better understanding of mental set and its negative effect on problem solving via their attempts to find facilitating clues. The clues inevitably lead to some in-class “Aha!” insight experiences that students cannot suppress. In addition to student clues, provide...
Table 1. The Four Series Problems Used in the Activity

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>OTTFSS__</td>
</tr>
<tr>
<td>(2)</td>
<td>EOREXN__</td>
</tr>
<tr>
<td>(3)</td>
<td>MΩ8M__</td>
</tr>
<tr>
<td>(4)</td>
<td>∞OΠΠΠF__</td>
</tr>
</tbody>
</table>

one final clue that all of the problems have essentially the same answer. Following a brief period to allow one more opportunity for solutions, solicit answers and carefully relate the answers to the mental sets involved for each series problem.

The solutions to all four problems involve the simple digit series (1, 2, 3, 4, 5, 6, 7, 8, …), but each requires overcoming mental set in a different way. For Series 1, the structure is that each alphabetic character represents the first letter of the successive digit words (O for ONE, T for TWO, …) so the answer is E for EIGHT. Based on one’s past experiences with series problems, the mental set involves viewing the alphabetic characters as single entities (letters) and not as parts (the first letters) of some related larger entities (the digit words). For Series 2, the structure is that each alphabetic character represents the last letter of the successive digit words (E for ONE, O for TWO, …) so the answer is T for EIGHT. The mental set created by past experience is the same as in Series 1 but further compounded by the normal ordering of words by their first letters and not their last. In addition, students who have solved Series 1 may become victims of the mental set of using the solution strategy for Series 1 in their attempts to solve Series 2. They may try to find a series of words in which the letters in Series 2 are the first letters of words.

For Series 3, the structure is that each of the successive digits is paired with its mirror image so the answer is 5 paired with its mirror image, 5. The relevant mental set in this case involves perceiving each character at a holistic level (e.g., seeing the third character in the sequence as the number 8). You can relate this difficulty to previous material in the introductory course on sensation and perception, specifically on pattern recognition, in which the brain automatically pieces the parts of the patterns together to interpret their meaning at the holistic level. Consciously dividing the characters into parts and thus overriding this automatic processing via past experience is a very difficult endeavor. For Series 4, the structures of Series 1 and 3 are combined so the answer is F combined with its mirror image, F. Thus, the interfering past experiences are the same as those for Series 1 and 3 combined. However, because many students with answers to Series 1 and 3 do not readily solve Series 4, this series reinforces the strong negative influence of the mental set created by these past experiences.

Discussion

This activity is appropriate for use in any course that includes coverage of thinking (e.g., cognitive psychology). It is not limited by class size. I have used variations of it in the introductory course, with class size ranging from 10 to 20 up to 200 to 300 students, and in cognitive psychology with intermediate enrollment sizes.

This term in my introductory class (N = 24), I surveyed my students about their engagement in the activity and their perceived value of the activity with respect to understanding mental set and the difficulty it creates in problem solving. For both aspects, the responses were unanimously positive. These data coincide with those for more informal assessments I have conducted in past terms. In brief, students both enjoy the activity and find it instructive.

To enhance the activity’s value, you should relate the difficulties created by mental set to the more encompassing concept of mindfulness using Matlin’s (1999, p. 263) discussion as a guide. By doing so, you will help the students to generalize their knowledge about mental set as a hindrance to problem solving. Be sure to contrast mindfulness with mindlessness. A more mindful approach allows for better problem solving because one is open to new approaches and aware of the possible relevance of multiple perspectives in a particular situation. Langer (1989, 1997) provided very good resource material for this discussion. In these two brief books, she not only described excellent examples of mindlessness that impact everyday thinking but also instructions for combating it with more mindful behavior.

Regardless of how much you extend this discussion, make sure to include this brief finish to the activity. Although your students will have developed a specific mental set for 1, 2, 3, 4, … solution strategies, general mental set will hinder their solution to this final problem. Have your class do the following series problem: 2HES__. What’s the answer? View this series in a mirror. Aha!

References

What Are Students Telling Their Friends? Teaching Responses to Lay Psychopathology Questions

Cynthia L. S. Pury
Clemson University

Undergraduate abnormal psychology students are likely to know people with mental disorders and thus may apply the material they learn in class to their friends and family. However, typical abnormal psychology courses offer students little to no instruction on how to apply course material to their real-life experience with mental illness. I present a technique to teach students to respond appropriately to questions about mental illness from others by using small-group-generated lay questions and appropriate answers followed by an individual oral examination. Student ratings suggest that the technique is useful in teaching them how to respond to lay questions.

Students enrolled in abnormal psychology courses are likely to know someone with a mental disorder. Mental illnesses affect 20% of Americans during their lifetimes (U.S. Department of Health and Human Services, 1999). Connor-Greene (2001) found 94% of students enrolled in an abnormal psychology course knew at least one person with a mental illness, with a mode of four people. These figures suggest that students have ample opportunities to apply class material to friends and family. Whether they apply the material appropriately is a question I propose can be partially addressed by a lay question–answer activity and related oral examination.

This activity grew out of my personal undergraduate experience when a friend asked me about his anxiety disorder. Despite my great interest in anxiety disorders—and my A in abnormal psychology—I was baffled. Many of my students report similar encounters followed by their own similarly bewildered responses. Although many psychology courses have an oral communication component, such activities usually focus on professional or research communication (e.g., Klugh, 1983; Moeller, 1985; Nadelman, 1990) or on development of counseling skills (e.g., Jackson, 1984). Few communication activities show students how to apply material learned in abnormal psychology to daily life (see Nolen-Hoeksema, 2001, for an exception).

I developed the following technique to give students direct practice applying course material to lay questions. Students work in small groups throughout the semester, writing questions lay people might ask about the current topic, then answering those questions. I use the best question–answer pairs in an individual oral examination at the end of the semester. The small-group work and the oral exam are worth 5% each toward the course grade.

At the end of the first class, students list past psychology courses, majors, and career interests. I then create small groups with varied experience and interests. I use groups of 5 or 6 students to ensure a critical mass for good discussion (a minimum of 3) when students are absent or (occasionally) drop the course. For my typical enrollment of 25 to 28 students, I have five groups per section. At the end of each major topic (50% to 75% of class periods), students break into their small groups. Before the first assignment, I instruct students in basic active listening (e.g., Moursund, 1985), have them use active listening to get to know another group member, then discuss the experience in class. This first, ungraded assignment takes 20 to 30 min.

For the remainder of the semester, each group writes a question a lay person might ask about the current major topic, along with an appropriate answer. My observation of these groups suggests that they generate meaningful questions—in fact, many questions come directly from students’ experiences. This activity usually takes about 10 min at the end of class, while I remain to answer questions. I then grade the written question–answer pairs, correcting any misperceptions or inappropriate responses. I modify any answer that goes beyond an undergraduate level of knowledge (e.g., “I think you should take some Prozac,” “You have panic disorder”) to something appropriate for undergraduate expertise (e.g., “There are drugs that can treat depression,” “I’m not an expert, but it sounds like it might be panic disorder”). I strongly encourage students to refer the questioner to a mental health professional if appropriate. I return photocopies of the corrected sheets to each student at the next class period, at which time I also discuss any serious or common misconceptions.

The best question–answer pairs become possible oral exam questions. At the end of the semester, students receive photocopies of these pairs. Following a problem method approach (McBurney, 1995), I emphasize that although the answers on the photocopies are good ones, other answers may be equally acceptable. Students then sign up for 3-min individual appointments with me. In larger enrollment courses, this duty could be divided among teaching assistants. During the appointment, the student draws a folded-up question out of a basket and hands it to me. I ask the question and grade the student’s response on a 0- to 2-point system. Zero points is failing, reserved for harmful answers such as providing seriously incorrect information, belittling the questioner, or pretending more expertise than appropriate. One point is passing, easily obtained by simple active listening and referral to a mental health professional. Two points is good, reserved for answers that are helpful, include active listening, are appropriate for students’ level of knowledge, and demonstrate a knowledge of course material. The majority of students earn 2 points, with a few earning 1, and no one yet earning 0. I purposely restrict the upper end of this scale to minimize both anxiety and incentives for students to overlearn answers to
dent reported being home watching television when her mother informed her of Princess Diana’s death. In the follow-up probe, the student reported being at her boyfriend’s house studying for an exam when he came home with the news. In both of these accounts, the student reported location, ongoing activity, and informant information. Aftermath disappeared in the second probe, and the other three categories of information disagreed.

The project led to lively discussions. Students were pleased and eager to describe their consistent memories as well as their memory discrepancies. One of the most useful aspects of the project is that it provides the class with self-generated examples of many of the concepts covered in an introductory course. For instance, when we want to discuss omission or elaboration, we ask students to search their probes for examples of accounts in which they preserved the gist but added or dropped details. When we discuss schemas, we ask students to search for inconsistent accounts and to speculate on how they may have used an existing schema to reorganize their memories. When examining the effects of emotion, we ask students who had experienced high levels of emotion at the time of encoding to compare their memories with students who had reported little emotion. The same type of discussion can take place when introducing rehearsal. We have found this exercise to be particularly successful for generating examples and illustrating the roles of emotion, rehearsal, consolidation, elaboration, omission, reconstruction, and schemas.

We assessed the effectiveness of this project in two ways: (a) by examining the impact on classroom test scores and (b) by examining course evaluations. We compared scores on examination questions concerning episodic and flashbulb memories ($M = 67.84$, $SD = 22.97$) to a sample of factual and conceptual questions concerning other memory phenomena on the same examinations ($M = 63.71$, $SD = 18.09$). Students scored higher on the episodic and flashbulb memory questions, $t(152) = 2.20$, $p = .03$. In contrast, students enrolled in sections that did not use this project tallied similar scores on these two sets of questions ($M = 63.44$, $SD = 18.34$ vs. $M = 63.38$, $SD = 21.97$), $t(241) = 0.03$, $p = .97$. We also compared scores on the episodic and flashbulb memory questions to performance on those same questions in course sections that did not use this project. Students who had completed this project ($M = 67.84$, $SD = 22.97$) scored higher on those questions ($M = 63.38$, $SD = 21.97$), $t(393) = 1.98$, $p < .05$. Thus, although the effect size is small, it does appear that recording their memories helped the students learn about episodic and flashbulb memories.

We solicited evaluations of the project and the course from students at the end of the semester. Students rated the course as 4.4 and the project as 4.6 on a 5-point scale ranging from 1 (poor) to 5 (outstanding). Written comments were also uniformly positive. Students described the project as “intriguing” and “interesting.” They were especially pleased to be “able to analyze ‘real data.’” Students also reported that they believed the project helped them to better understand memory processes.

Overall, we believe this project is rewarding and instructive for both faculty and students. The project improved student involvement and learning. It required just a few minutes of class time to collect the memory probes and as much time in discussion as the instructor desires. It is a project that can be performed every semester. Many events occur that might trigger vivid memories. For example, recent flashbulb memory research (for a review, see Conway, 1995) has examined memories for the assassination attempt on President Reagan, the San Francisco earthquake, and the Los Angeles earthquake, among others. Although events such as these do not occur at the onset of every semester, we have been able to use the same event for several semesters until another unexpected and emotional event occurs. Finally, we believe that this project is flexible enough to be useful in a number of courses. In introductory psychology we have used it as a vehicle for discussion of memory phenomena. We have also used this project in upper level learning and memory courses. Here, we require more extensive data analysis, a literature review, and completion of a laboratory paper. As a result we have been able to impart concepts in research design and analysis as well as teach about memory phenomena.

References


Note

Send correspondence to Mark Sudlow Hoyert, Psychology Department, Indiana University Northwest, Gary, IN 46408; e-mail: mhoyert@iunhaw1.iun.indiana.edu.

Demonstrating the Concept of Illusory Correlation

Jay W. Jackson
Indiana University–Purdue University

The concept of illusory correlation helps explain how some social stereotypes are partly due to basic cognitive processes. I describe a classroom activity that essentially replicates, in a simplified format, a classic study of the illusory correlation effect. In addition to effectively demonstrating the phenomenon of illusory correlation, this activity provides a stimulus for discussing related social psychological issues and methodological procedures.

Research has demonstrated that people pay special attention to distinctive or rare events and recall such events with
I have conducted a classroom demonstration of the illusory correlation effect that "replicates" a classic experiment conducted by Hamilton and Gifford (1976, Study 1). Students read about people from a numerical majority group (Group A) and a numerical minority group (Group B). Both groups are associated with more desirable behaviors than undesirable behaviors, and the proportion of desirable and undesirable behaviors is the same for both groups (a 9:4 ratio). Because both Group B and undesirable behaviors are rare events, students tend to form illusory correlations and judge Group B less favorably than Group A. Just as researchers have found the effect to be highly reliable across various laboratory settings (Hamilton & Sherman, 1996; Lieberman, 1999), I have found the results to be quite reliable in the classroom. This demonstration helps illustrate the role that basic cognitive processes play in the development of social stereotypes.

The demonstration takes about 20 min and requires a total of 39 slides, each containing a description of a Group A member or Group B member engaging in either a desirable or an undesirable behavior (see Table 1). The descriptions may be presented using computerized slides, overheads, traditional slides, index cards, or other means of presenting written visual stimuli. Reading the descriptions also works as well.

Following the procedures used by Hamilton and Gifford (1976, Study 1), I inform the students that they will see a series of statements, each describing a person performing some type of behavior, and explain that each person described will belong to one of two groups, simply referred to as Group A and Group B. I ask students to read (or listen to) each statement carefully as it is presented. I then present each statement, in a random order, for about 8 sec. After they have seen all the statements, the students complete a rating form. On this form, students indicate their "first impressions" by judging each group, on 7-point Likert scales ranging from 1 (strongly disagree) to 7 (strongly agree), as being generally popular, lazy, unhappy, intelligent, honest, irresponsible, helpful, and unpopular. Second, the students provide frequency estimates regarding the behaviors exhibited by the members of each group (from Hamilton & Gifford, 1976, Study 1):

There were 26 statements about Group A members. In your estimate, how many of these 26 statements described
Disagree) to 7 (strongly agree).

Illusory correlation may help, in part, explain this tendency because African Americans constitute a minority and undesirable behaviors are uncommon events.

After the demonstration, I ask the class what other stereotypes might be explained by this concept (characteristics of Asian Americans, religious leaders, the mentally ill, and adolescents are common responses) and how the mass media contribute to these phenomena. For example, when a mentally ill person commits an act of violence (e.g., the shooting of John Lennon or Ronald Reagan), the person's mental condition commands attention, as does the violent act itself. Both are relatively infrequently experienced by the general public, and thus their co-occurrence is especially memorable.

In discussing the possibility that the results were due to negative feelings or beliefs about minorities in general, I point out that Hamilton and Gifford (1976, Study 2) conducted a second study in which the rare behaviors were desirable, rather than undesirable. This change resulted in the minority group (Group B) being rated more favorably on the traits than the majority group.

Of course, the predicted results will not always be obtained for each item and not every student will express the expected bias. Such contrary results provide an opportunity to discuss the probabilistic nature of psychological research and the importance of individual differences. I usually ask the students to discuss why some people might be more prone to the effect than others.

To highlight a methodological issue, I also ask the class why Hamilton and Gifford (1976) used “Group A” and “Group B” in the experiments instead of actual groups. Of course, the reason is to control for previously established associations between groups and traits. Unlike those who judged Groups A and B, people often have preexisting biases for or against real groups. Further research (e.g., Hamilton, Dugan, & Trolier, 1985; Hamilton & Rose, 1980) has shown that preexisting stereotypes can lead people to see correlations that are not there (e.g., doctors who are wealthy, accountants who are timid, librarians who are quiet). Information that is consistent with stereotypes is easier to encode and is more likely to be recalled later (Rothbart, Evans, & Fulero, 1979). Students generally find this demonstration interesting and valuable in clarifying the concept of illusory correlation. Student evaluations of two recent demonstrations appear in Table 3. In addition to clarifying a cognitive basis of stereotyping

### Table 2. Mean Trait Ratings and Frequency Estimates for Two Demonstrations of the Illusory Correlation Effect

<table>
<thead>
<tr>
<th>Trait</th>
<th>Session 1</th>
<th></th>
<th>Session 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
<td>Group A</td>
<td>Group B</td>
</tr>
<tr>
<td>Popular</td>
<td>4.64</td>
<td>3.86</td>
<td>4.12</td>
<td>4.33</td>
</tr>
<tr>
<td>Lazy</td>
<td>2.28</td>
<td>3.41</td>
<td>3.16</td>
<td>4.20</td>
</tr>
<tr>
<td>Unhappy</td>
<td>3.42</td>
<td>4.94</td>
<td>3.54</td>
<td>3.83</td>
</tr>
<tr>
<td>Intelligent</td>
<td>5.57</td>
<td>4.14</td>
<td>5.00</td>
<td>4.12</td>
</tr>
<tr>
<td>Honest</td>
<td>5.35</td>
<td>3.00</td>
<td>4.00</td>
<td>3.78</td>
</tr>
<tr>
<td>Irresponsible</td>
<td>2.65</td>
<td>5.28</td>
<td>3.86</td>
<td>4.13</td>
</tr>
<tr>
<td>Helpful</td>
<td>5.17</td>
<td>3.41</td>
<td>4.87</td>
<td>4.39</td>
</tr>
<tr>
<td>Unpopular</td>
<td>3.07</td>
<td>4.21</td>
<td>3.39</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Note. Trait ratings were based on a scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Item MS D

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I enjoyed this demonstration.</td>
<td>5.78</td>
<td>0.69</td>
</tr>
<tr>
<td>This demonstration helped clarify the concept of illusory correlation.</td>
<td>6.00</td>
<td>0.39</td>
</tr>
<tr>
<td>I would recommend that this demonstration be used in future classes.</td>
<td>5.92</td>
<td>0.47</td>
</tr>
<tr>
<td>This demonstration was interesting.</td>
<td>5.92</td>
<td>0.61</td>
</tr>
<tr>
<td>I will probably tell my friends or family members about this demonstration.</td>
<td>5.42</td>
<td>1.01</td>
</tr>
<tr>
<td>After participating this demonstration, I want to learn more about the nature of stereotypes.</td>
<td>5.36</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Note. N = 37. Responses were based on a scale ranging from 1 (strongly disagree) to 7 (strongly agree).
and prejudice, this demonstration stimulates students to think about other factors that contribute to how people judge individuals and groups.

References


Notes

1. Many thanks to David Hamilton for kindly sharing his research materials and to the anonymous reviewers for their helpful comments on an earlier draft of this work.

2. Send correspondence to Jay W. Jackson, Department of Psychology, Indiana–Purdue University, 2101 Coliseum Boulevard, Fort Wayne, IN 46805–6403; e-mail: jacksonj@ipfw.edu.

"Parenting" Students: Applying Developmental Psychology to the College Classroom

Mary Barnas
Marietta College

In this article, I use Baumrind’s (1971) research on parenting styles to help understand the structure and dynamics of the college classroom. I argue that the way professors view their students affects course syllabi and teaching style. The discussion also reflects the development of my own teaching style and philosophy.

After 6 years, my teaching philosophy and attitude toward my students completely changed. I realized that my years of training as a developmental psychologist were applicable to my teaching. Underlying the changes in my classroom poli-

The Permissive Teacher: “Come to Class if You Want To”

I began teaching at a private, liberal arts institution. I was committed to the notion that my students were adults. I had no attendance policy and no “policing” strategies in place (e.g., pop quizzes) to check whether students kept up with their reading. External contingencies would do little to help them after graduation. Quickly, I became frustrated. Attendance was poor and only a few students participated in the class. After 3 years, I accepted a new position at a somewhat smaller institution where teaching is the primary mission. However, nothing had changed. Attendance and participation were low, and my enthusiasm for this noble profession was reaching a similar level. Still, I held fast to my values. It was only after the students began to ask for the very policies I assumed they resented that I started contemplating fundamental changes. I was shocked to read comments on my evaluations suggesting that I should “have an attendance policy so we will come to class” and “have quizzes so that we won’t put off all of the reading.” One day, I uncharacteristically called on a student in my class and kept questioning him until he at least made an attempt to respond. After class, the student approached me and said. “Dr. Barnas, thanks for doing that. That’s the only way I’m going to get into this class and start really thinking.” The final indication that something may have been wrong with my teaching philosophy came when a student I knew fairly well confided in me that some students were under the impression that I did not care about their investment in my courses because I was lenient about deadlines for assignments.

I began to examine my teaching philosophy and to explore the relation between the developmental literature on parenting styles and my behavior thus far in the classroom. I had been exhibiting a permissive style of teaching. I had few rules and rarely enforced those that existed, not because I did not care, but because I believed the students would be best served by making these decisions for themselves. Baumrind and Black (1967) found that children parented under the permissive style are often unsure of themselves and lack self-control. It seems as though allowing self-imposed limits on immature human beings has the opposite of the intended effect. They flounder and resent authority figures for not providing guidelines. Perhaps although physically mature, these students were not ready to face the adult world of self-imposed deadlines. What next?

Authoritarian Teaching: “Come to Class or Else”

I changed my course policies starting with attendance and then adding pop quizzes. I would not accept late papers. I gave no explanation for these policies but only made it clear I would enforce them. The result was that students were attending my classes in record numbers, and they were more of-
Faculty Forum
Mark Sudlow Hoyert, Cynthia D. O'Dell, Jay W. Jackson, Mary Barnas, Margaret M. Nauta, William Buskist, Steven A. Meyers & Loreto R. Prieto
Available online: 20 Nov 2009

To cite this article: Mark Sudlow Hoyert, Cynthia D. O'Dell, Jay W. Jackson, Mary Barnas, Margaret M. Nauta, William Buskist, Steven A. Meyers & Loreto R. Prieto (2000): Faculty Forum, Teaching of Psychology, 27:4, 272-284
To link to this article: http://dx.doi.org/10.1207/S15328023TOP2704_06

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
the alphabet listed vertically in random order. Beside each letter is, first, a rating scale, ranging from 1 (do not like) to 5 (like a lot), and then four blank lines under the column headings IYFN, NIYFN, IYLN, and NIYLN. Acknowledging that the task may seem rather silly, we ask students to rate rapidly how much they like each letter. Once all students have finished, we ask them to print their first and last names at the top of the handout. Then we explain the letters above the columns: IYFN stands for In Your First Name; NIYFN, for Not In Your First Name; IYLN, for In Your Last Name; and NIYLN, for Not In Your Last Name. Next, we ask the class to fill in their letter ratings under the appropriate columns. For example, if a student’s first name was Kevin, he would fill in his ratings for K, E, V, I, and N in the IYFN column. He would fill in his ratings for all other letters in the NIYFN column. If this student’s last name was Arnold, he would fill in his ratings for A, R, N, O, L, and D in the IYLN column and his ratings for all other letters in the NIYLN column. When all students have filled ratings in the appropriate columns, we ask them to calculate their mean for each column. After they complete this task, we ask the following question: “How many of you had a higher average for letters in your first name than for letters not in your first name?” The majority of hands usually go up. We repeat the question for last name with, typically, the same result.

At this point, we explain the name letter effect and mere ownership effect, plus the supportive research noted previously. Instructors tying the topic into a discussion of other self-serving biases could conduct demonstrations of those on the same day. Alternatively, instructors could use the name letter effect as an example of processing without conscious awareness.

Evaluation

We have found this technique to be successful on multiple occasions. Twice after conducting the demonstration in introductory psychology, we collected and double-checked students’ handouts and then calculated group means. In one class, in which 45 students participated, mean liking for letters in students’ first names was 3.7, and mean liking for letters not in their first names was 3.0. Mean liking for letters in their last names was 3.5 and for letters not in their last names, 3.0. In the other class, means for letters in and not in students’ first names were 3.8 and 3.0, respectively, and means for letters in and not in their last names were 3.6 and 3.0, respectively. Within-group t tests showed that all differences were significant at the p < .0001 level. In the first class, we also found that those present for the demonstration were more likely to answer a multiple-choice test question on the name letter effect correctly, $\chi^2(1, N = 80) = 35.7, p < .0001$, than those who were absent. Too few students were absent on the day of the demonstration in the second class to conduct a similar analysis although, of those present, over 80% successfully answered the test question.

In our experience, this is a simple and reliable demonstration, one that is useful for social, cognitive as well as introductory courses. It is also an interesting demonstration, as students are amused to learn of their latent letter preferences.

References


Note

Send correspondence and requests for copies of the handout to Angela Lipsitz, Department of Psychology, Northern Kentucky University, Highland Heights, KY 41099; e-mail: lipsitz@nku.edu.

“Me Conform? No Way”: Classroom Demonstrations for Sensitizing Students to Their Conformity

C. R. Snyder

University of Kansas, Lawrence

This article presents 9 class demonstrations for increasing student awareness of personal conformity. One demonstration uses the Milgram (1965) obedience studies, a second and third involve self-fulfilling prophecies, a fourth and fifth illustrate the role of clothing, a sixth and seventh focus on peer pressure, an eighth examines individual differences in conformity, and a ninth provides a humorous example of conformity.
When I began teaching about conformity, I was startled at how college students denied engaging in such behaviors. How could students see themselves as being so different from others who do conform in psychology experiments and in everyday life? I believed that I needed to teach about conformity so that my students would see it as something that they might do. A deeper understanding of this important phenomenon should result from such personal insights about their conforming. In this article are nine demonstrations I have developed for sensitizing students to their own conformity.

**Demonstration 1: A Twist on Milgram Obedience**

In that portion of my syllabus dealing with conformity, after the first lecture named *Obedience*, the following phrase appears: “Bring an Empty Soda Can to Class!” When that lecture day arrives, I show the film depicting the classic Milgram (1965, 1974) experiment in which the research participants increasingly deliver higher electric shocks to another person when directed to do so by an experimenter. My students become very tense, laugh nervously, and exchange uneasy glances while watching this film. After the film, I highlight the findings about how upwards of 50% to 60% of the research participants manifested the highest level of obedience in delivering the electric shock. I ask students who brought an empty soda can to place that can in their left hand. Next, I ask my students to raise their right hand if they were absolutely certain that, in the same experimental situation, they would not shock the other person to the highest 450-volt level. Virtually all students raise their right hands. I then ask everyone to raise his or her left hand, whereupon a large number of the students are holding up empty soda cans. After about 30 sec, I ask, “Why do you have those soda cans in your left hands?” The students look at each other, and eventually one says, “You told us to bring the cans.”

There is yet more quiet on my part as this lesson sinks in. Eventually, a student protests by asserting, “There is no similarity between shocking another person and bringing a soda can to class!” By now, my students’ interest is highly aroused.

**Demonstration 2: Rosenthal’s Classroom “Spurters”**

I begin this demonstration by explaining the specifics of the Rosenthal and Jacobson (1968) study in which one group of grade school children was labeled *spurters* (i.e., on the verge of an academic intellectual increase), and the other group of children received no such label (the *nonspurters*). I then feign having noticed that my “brighter” students—my spurters—seem to sit on the right side of the auditorium (or the left side because it makes no difference for purposes of the demonstration). I move my podium in front of these bright students, along with the screens with the overheads so that these “good” students can have easy viewing. All the while, I pay attention only to their reactions to being treated as dumb. Likewise, some of the spurters report that they found themselves reveling in their special treatment. To their surprise, several students admit to getting caught up in the demonstration.

**Demonstration 3: Blue Eyes/Brown Eyes**

In 1968, an Iowa grade school teacher named Jane Elliott performed a classroom demonstration (see Peters, 1987) in which she gave preferential treatment to the brown-eyed children one day and treated the blue-eyed children as if they were stupid and irresponsible. On the second day, she reversed the preferential treatment such that the blue-eyed children were the favored ones. My students watch the videotaped version of this experiment (http://www.newsreel.org/films/blueeyed.htm) with riveted attention.

I close this Blue Eyes/Brown Eyes demonstration with an in-class example of obedience. Namely, 5 min before the end of the lecture period, I ask the blue-eyed students to stand up—which they do. I then tell them to leave—which they do. The brown-eyed students are sitting at their desks, looking perplexed as I gather up my materials. Finally, a brown-eyed student pipes up, “Hey, what is going on here?” My response is, “It looks like obedience.” The students and I enjoy a good laugh, and we discuss this incident fully at the beginning of the next lecture period.

**Demonstrations 4 and 5: Clothes Make the Conformist**

I have two favorite conformity demonstrations using clothing. In the first one, I ask a group of about 10 students to come up in front of the class. Standing in line, it becomes apparent that the majority of students are wearing “the uniform” (a tee shirt and blue jeans). The discussion that ensues on the part of students as to why this behavior is not conformity is hilarious and creative in content.

My other clothing demonstration involves the colors that we wear. The female color, from the time of first being wrapped in terry cloth at the hospital, is pink. Men, on the other hand, are dressed in blue from day one onward. To drive home the results of this exercise, I use a $2 \times 3$ ($\text{Female/Male} \times \text{Pink/Blue/Neither}$) matrix on an overhead. I then ask all of the women to stand and count those who are wearing a pink shirt—that number goes in the appropriate box in the matrix (i.e., female, pink). Next, I count the women who are wearing a blue shirt, and I place that number in the appropriate box. Last, I count the women who are wearing a shirt that is neither pink nor blue and record that number in the matrix. I follow the same routine for the men.

The women wear more pink than the men, and about the same amount of pink as blue. Conversely, the men wear more blue than the women, and they rarely wear pink (i.e., over-
whelmingly more blue than pink). For the neither pink nor blue categories, the women and men are basically the same. This demonstration shows that color sex conformity is alive and well on college campuses today.

Demonstrations 6 and 7: “Everyone Is Doing It”

I set this demonstration as a bar scene, and I lower the lights and have country and western music (or other types) playing. There are two stools at the front of the classroom, and I invite a woman and man to participate in a “pickup drama.” I ask the man to make repeated romantic overtures to the woman. What he says and how she responds are not scripted. This exercise usually ends with the man using verbal coercion such as “everyone is doing it.” The female students report that this demonstration is quite accurate, and they point out that they may hear a similar logic when they have agreed to have sex, but say “no” to unprotected intercourse (i.e., the man not wearing a condom).

In the ensuing discussion, some of the male students suggest that women who say “no” do not necessarily mean it and that this all part of “the game.” Because of such male reactions, I have added a second demonstration in which I ask for one male and three female volunteers. The role of the women is to get the man to try cocaine. I instruct the women to use the everyone is doing it coercion line. Additionally, I tell the women to interpret any resistance on the part of the man as being phony—because he “really wants to try it.” The man is asked merely to continue saying “no.” The women verbally hammer him to “not be wimp and to be part of the cool crowd.” The man says “no,” but the women are unrelenting, and then use the “Oh honey, we know that you don’t really mean ‘no’!” This latter demonstration helps male college students to see that they should take a no at face value.

Demonstration 8: The Uniqueness Seekers As Nonconformers

Some people are less likely to conform because of their need for uniqueness. Uniqueness needs represent the desire to be different from others—a willingness to stand out (Snyder & Fromkin, 1980). Research has shown that high- as compared to low-need-for-uniqueness persons are less likely to conform. To give students an index of their need for uniqueness, I administer the 32-item Need for Uniqueness Scale (Snyder & Fromkin, 1977). I next give the norms for the scale, and students can gain insights into their likelihood of conforming.

Demonstration 9: The Student Becomes Instructor … A Little Fun

I end this conformity section by planning it such that only about 10 min remain the class. At that point, I ask if any student would like to come to the podium and make a comment about or give a final example of conformity. The students are abuzz. Finally, a student strides down to the podium and, with a big smile, announces loudly, “Class dismissed!” The class quickly exits, conforming to the command of their new “authority.”

Do These Demonstrations Work?

I routinely ask students at the end of my course to give anonymous feedback about the most important things they have learned. Overwhelmingly, students list these conformity demonstrations as having made them more sensitized to the circumstances when they are conforming as well as helping them to become more resistant to conforming. For female college students, the most common scenario has involved resisting being coerced into having sex (or unprotected sex). Men describe how they have fought peer pressure to take drugs. Additionally, when I meet former students after they have graduated (e.g., at airports, reunions), they recall one or more of the conformity demonstrations and ask whether I still am doing them. These former students also recount how they applied the lessons of these conformity exercises in actual personal circumstances.

The conformity forces are numerous and strong for college students today. I believe that these demonstrations experientially help them to understand and to cope with the pressures of such conformity.

References


Note

Send correspondence and requests for more information about the Need for Uniqueness Scale to C. R. Snyder, 1415 Jayhawk Boulevard, 340 Fraser Hall, Graduate Training Program in Clinical Psychology, Department of Psychology, University of Kansas, Lawrence, KS 66045; e-mail: crsnyder@ku.edu.
Faculty Forum


Available online: 20 Nov 2009


To link to this article: http://dx.doi.org/10.1207/S15328023TOP3001_11
or attitude measure was composed of Likert-type questions in the end-of-semester student evaluation form.

Results

The average number of PCQ assignments submitted by the JiTT group was 11.24 of 12 (SD = 1.45), with a minimum value of 7 and a maximum value of 12. At the end of the semester, we asked the students in the JiTT group to evaluate their perceptions of the efficacy of the PCQs for learning and their enjoyment working with the approach. They responded on 5-point Likert rating scales. Table 1 shows the percentage of students who agreed or strongly agreed with each of the questions. In general, they were very positive about the use of this approach, particularly about discussing student answers in class. When we compared the final exam scores (possible range from 0 to 100), the JiTT class (M = 76.25, SD = 11.07) performed better than students in the control class (M = 72.39, SD = 8.89), t(119) = 2.13, p = .04, two-tailed, d = .38.

Discussion

The results supported our hypothesis that the JiTT students would like the approach and would find it facilitated their learning of statistics. Perhaps they were motivated to submit their best work because of the points they received and of the possibility that the instructor might show their responses to the class.

The higher final exam score for the JiTT group provided some empirical support for our second hypothesis. However, given the quasi-experimental nature of the design and the potential problem of experimenter bias, these results should be considered as suggestive rather than definitive.

Several drawbacks occur with the JiTT approach. First, instructors need to learn how to create Web forms and have a server to store them. Second, the approach takes some additional time to implement. This instructor spent an extra 1 to 2 hr per week preparing for class. Third, because these assignments were not due in class, some students complained about forgetting or missing the Internet submission deadline of 8:00 a.m. on Friday. Lastly, some students did not have convenient access to the Web.

The JiTT approach seemed to facilitate the teaching of statistics. We have also used it successfully in general psychology and research methodology classes. In general, students like using this approach in each of these classes. There is a question of whether this approach might be more effective in lower division or upper division courses that needs to be investigated. In conclusion, JiTT created a classroom experience that was meaningful, timely, and effective for both the instructor and the student. It provided immediate feedback to students about their level of understanding, and students liked using the method.

Table 1. Percentage Agreed or Strongly Agree to Questions That Evaluated the JiTT Method

<table>
<thead>
<tr>
<th>Questions</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCQs were extremely effective for learning statistics</td>
<td>69</td>
</tr>
<tr>
<td>It was helpful for the professor to discuss answers in class</td>
<td>90</td>
</tr>
<tr>
<td>PCQs facilitated my understanding of statistical concepts</td>
<td>71</td>
</tr>
<tr>
<td>I enjoyed the PCQs because they were on the Web</td>
<td>56</td>
</tr>
<tr>
<td>I liked the PCQs because I could do them on my own time</td>
<td>67</td>
</tr>
<tr>
<td>PCQs encouraged me to read ahead</td>
<td>48</td>
</tr>
<tr>
<td>PCQs enhanced communication between student and professor</td>
<td>50</td>
</tr>
<tr>
<td>PCQs should be used in future statistics classes</td>
<td>77</td>
</tr>
</tbody>
</table>

Note. N = 55. We removed one student from the JiTT group because this student completed only 4 of the 12 PCQs during the semester. We also removed a student matched on the basis of the final exam score from the control group. JiTT = Just-in-Time Teaching; PCQ = preclass questions.

References


Note

Send correspondence to James O. Benedict, Department of Psychology, James Madison University, Harrisonburg, VA 22807; e-mail: benedicto@jmu.edu.

Using a “New Classic” Film to Teach About Stereotyping and Prejudice

Andrew N. Christopher and Jamie L. Walter

*Albion College*

Pam Marek

*Anderson College*

Cynthia S. Koenig

*St. Mary’s College of Maryland*

We describe a method for helping students learn about stereotype formation and prejudice by having them watch and discuss characters and scenes in the movie *The Breakfast Club* (Tanen &
Hughes, 1985). We identify clips that address specific principles, notably the contact hypothesis. Students’ positive evaluations of this method suggest the film is an effective way to teach about stereotyping and prejudice.

Current and classic films can be useful pedagogical devices to teach a variety of topics in psychology. For example, Conner (1996) used Total Recall to teach concepts in cognitive psychology. Hemenover, Caster, and Mizumoto (1999) used Ordinary People to facilitate understanding of Carl Rogers’s concept of unconditional positive regard, and they used films such as One Flew Over the Cuckoo’s Nest, The Fisher King, and Regarding Henry to illustrate a variety of mental disorders in written assignments. The variety of ways in which teachers can use films to enhance the learning environment appears to be limitless (see also Anderson, 1992; Boyatzis, 1994; Desforges, 1994; Richard, 1996).

Most recently, Roskos-Ewoldsen and Roskos-Ewoldsen (2001) provided a list of movie clips to stimulate discussion of traditional social psychological topics (e.g., person perception, obedience). We extend their contribution by specifying how teachers can use a feature film, The Breakfast Club (Tanen & Hughes, 1985), which depicts five high school students spending a Saturday together in detention, to teach about stereotyping, prejudice, and the reduction of prejudice, with an emphasis on the contact hypothesis (Allport, 1954; Cook, 1985). The film is suitable for use in a variety of psychology courses (e.g., social psychology, introductory psychology, group dynamics, stigma).

In his social psychology course, the first author spends approximately five 50-min class sessions on the topic of stereotyping and prejudice. Prior to seeing The Breakfast Club (Tanen & Hughes, 1985), students read pertinent information in their texts to prepare for class discussion of concepts illustrated in the film. Instructors may choose to show the entire movie or selected clips. Each of the five students in the film embodies elements of a stereotype, which the film depicts as an “athlete,” a “princess,” a “brain,” a “criminal,” and a “basketcase.” They began the day with seemingly nothing to talk about because they tended to see each other “in the simplest terms and most convenient definitions” (Brian, in Tanen & Hughes, 1985, 0:2:47).

**Concepts Related to Stereotyping, Prejudice, and Discrimination**

Table 1 contains example clips and quotes from The Breakfast Club. Teachers can use many clips to illustrate multiple concepts. Typically, discussion begins with an overview of the differences between stereotyping, prejudice, and discrimination and then turns to the causes and consequences of stereotyping and prejudice. Discussion concludes with an examination of ways to reduce these tendencies.

Reducing Stereotypes and Prejudice

In addition to illustrating stereotype and prejudice formation, The Breakfast Club provides a vehicle through which students can apply the conditions of the contact hypothesis (Allport, 1954; see also Cook, 1985). According to the contact hypothesis, there are five conditions of contact between group members that must be satisfied to blur stereotypes and reduce prejudice (see Cook, 1985, p. 453). Without meeting these conditions, it is unlikely that mere contact will be successful.

First, out-group members must possess traits and display behaviors that challenge the negative stereotypes of their groups. Clearly all five characters do so. Brian, the brain, smoked marijuana, failed an easy class, and carried a pornographic picture in his wallet. Andy, the athlete, allowed his father to control his life. Alison, the basketcase, actually does desire human contact. Claire, the princess, with the “perfect” life, smoked marijuana and had parents who fought a lot. John, the criminal, willingly took responsibility for all of the students skipping out of detention, thus helping dispel the belief that he was primarily self-interested. Indeed, of the five conditions of the contact hypothesis, it is perhaps this first condition that is best illustrated, at least as indicated by the proportion of class time spent discussing it.

Second, local authorities and norms should support the contact. All characters (except Alison) were required to be in detention. Mr. Vernon, the school principal, required all students to write essays about who they were. Likewise, his disparaging remarks were directed toward the students as a group, not at certain individual students. Such behaviors on Mr. Vernon’s part support the notion that the individual students were each part of the same larger group.

Third, contact should be among individuals of equal status. As soon as the characters assumed their seats for detention, Mr. Vernon made it known that no one was “special” in the daylong detention. Mr. Vernon made no attempt to treat anyone with particular respect, thus clearly establishing their “equal status” on this one day. Later, as the characters discussed a variety of issues, they each realized their unique potential. More importantly, the characters began to understand they were more alike than they thought.

The remaining conditions are that contact should occur at the individual level and that members of different groups work together toward a common goal. In detention, each student was a lone representative from his or her social group, and there was no opportunity for socializing with anyone other than fellow detainees. Thus, each encounter allowed students to approach each other as individuals rather than as members of a specific social group. As the movie progressed and the students began to perceive each other as individuals, they developed a single unifying goal, to make the principal aware of his stereotypical beliefs, prejudicial attitudes, and discriminatory behavior toward different groups of students. Brian’s letter illustrated this point rather eloquently:

> You see us as you want to see us, in the simplest terms and most convenient definitions. But what we found out is that each one of us is a brain, and an athlete, and a basketcase, a princess, and a criminal. Sincerely yours, The Breakfast Club. (Tanen & Hughes, 1985, 1:34:18)

**Evaluations and Conclusions**

At the end of each of five different semesters, students (N = 162) assessed use of the entire film by responding to five items...
on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Responses indicated that students found the movie enjoyable ($M = 5.79, SD = 1.10$), helpful in understanding the principles of stereotyping and prejudice ($M = 6.66, SD = 0.74$), easier to learn from than the textbook ($M = 6.14, SD = 1.01$), a good supplement to lecture ($M = 5.64, SD = 1.24$), and suitable for use in future classes ($M = 6.72, SD = 0.75$). Moreover, feedback suggested that *The Breakfast Club* also demonstrated the “ubiquitous” nature of stereotyping and prejudice (Aronson, Wilson, & Akert, 2002, p. 457).

Enthusiastic student feedback indicates that *The Breakfast Club* may be a useful tool for stimulating discussion of issues related to stereotyping and prejudice, thus complementing the film’s use in teaching theories of adolescent development (Desforges, 1994). Instructors who do not want to invest two class sessions in watching the entire film might show the suggested clips to illustrate specific principles. In such cases, it might be helpful for teachers to provide an overview of the scene and the characters. Because the movie is set in a seemingly all-White school, it may be important to discuss how the movie would differ with a more heterogeneous cast. Likewise, given current events, teachers may ask students to reflect on the stereotypes and prejudices that form between different religious groups. Indeed, the racially homogeneous cast only amplifies *The Breakfast Club*’s ability to stimulate generalizations to a range of stereotyping phenomena and to spark interest in techniques for reducing stereotyping and prejudice.

### References


---

1. We thank an anonymous reviewer for this suggestion.
Applied Animal Behavior Course: A Service-Learning Collaboration With the Humane Society

Lori R. Kogan  
Clinical Sciences  
Colorado State University

Julie A. Kellaway  
Colorado State University

Service-learning programs apply classroom principles to real-world situations and help communities by providing an often untapped resource of volunteers. This article describes how we applied the service-learning teaching approach to an undergraduate psychology animal behavior class in conjunction with a local humane society. Undergraduate psychology students learned operant conditioning techniques and applied this knowledge to 52 dogs housed at a local humane society. Students viewed the course as a positive experience and believed it offered them the opportunity to practice classroom knowledge in an applied setting. The article describes the course and the effects of this service-learning program on the community.

Academic service learning is a unique teaching and learning approach that integrates community service with academic study to enrich learning, teach civic responsibility, and strengthen communities (University of Colorado, 2000). Continuing to grow in popularity, service learning allows students to apply classroom material to community problems and promotes both personal and academic student goals (Eyler, 2002; Stukas, Clary, & Snyder, 1999).

Previous studies have demonstrated the success of service-learning courses in terms of both student approval (Chapdelaine & Chapman, 1999; Jurgens & Schweitzer, 2002) and community satisfaction (Hardy & Schaan, 2000; Holley, 2002). Research journals, however, have published few studies pertaining to service learning in undergraduate animal behavior psychology classes (Daniel & Perelle, 1987; Lukas, Marr, & Maple, 1998), and these studies have focused on zoo animals. Although animal training in zoos is necessary and beneficial, issues surrounding companion animal behavior problems resonate more intimately with many people. Human–dog relationships can be very rewarding, but behavior problems are the most frequent reason given by people who decide to relinquish their dogs and are thought to be responsible for 50% to 70% of all dogs euthanized (Miller, Staats, Partlo, & Rada, 1996; Salmon et al., 1998; Spencer, 1993).

Research pertaining to animals in humane societies has found that the stressful environment in animal shelters can exacerbate preexisting behavior problems, as well as create new ones, and even the best shelters create stress for dogs (Tuber et al., 1999). Human interaction, however, can reduce the behavioral and physiological effects of dogs’ stress (Fuller, 1967; Hennessy, Davis, Williams, Mellott, & Douglas, 1997; Hennessy, Williams, Miller, Douglas, & Voith, 1998). Unfortunately, few shelters have the extra personnel needed to interact with shelter dogs on a regular basis, and humane societies have traditionally relied on volunteers to fulfill this need. College students, and in particular, undergraduate psychology students, may provide a better solution. As Tuber et al. (1999) explained, “behavioral sciences [programs] are in a unique position to offer assistance to shelters in developing a supportive and positive environment, increasing the adoptability of pets, and maintaining adoptions” (p. 385). Tapping into this training arena for applied animal behavior experience, the first author developed and taught an undergraduate animal behavior service-learning course in collaboration with a local humane society.

Course Description

Undergraduate students at a large western university were offered a three-credit upper division psychology course titled Applied Animal Behavior in Spring 1999 as a complementary course to the more traditional, preexisting Psychology of Learning and Animal Behavior courses. The course consisted of weekly 50-min classroom sessions and twice-weekly 1-hr sessions at the local humane society. To assist with the supervision of training sessions, I selected two undergraduate seniors with past experience in animal behavior and dog training who received three credits for their involvement in the course.

I designed the course to help students become familiar with different types of obedience training and animal behavior through hands-on experience at the humane society, class speakers, and reading material; encourage students to think critically about domestic animal welfare issues (e.g., overpopulation, euthanasia); create an opportunity for students to apply “book knowledge” of animal behavior training principles to real-life situations; and forge a mutually beneficial working relationship between the University and the humane society. Additional course objectives were to help students identify and correct simple behavior problems, teach basic obedience commands, become familiar with the programs and operations of the local humane society, create op-
You Are What You Wear: An Interactive Demonstration of the Self-Fulfilling Prophecy

Michelle R. Hebl and Eden B. King
Rice University

This article presents evidence in support of an interactive classroom activity that demonstrates the self-fulfilling prophecy. After a brief description of the topic, 5 volunteers from an introductory psychology course of 81 students blindly donned one of several labeled hats (i.e., intelligent, attractive, good leader, annoying, lazy). The instructor told volunteers to treat each other in accordance with the labels while they completed a series of group tasks. During the task, volunteers acted in accordance with their label, thereby confirming the self-fulfilling prophecy. Findings indicated that the demonstration does elicit the self-fulfilling prophecy and that students’ understanding of this phenomenon improves. Furthermore, students overwhelmingly reported that this demonstration was interesting and enjoyable and recommended its use in future classes.

Although some psychological phenomena discussed in introductory and social psychology classes resonate easily with students’ perceptions of the world, other phenomena are more counterintuitive, abstract, and difficult to teach students of psychology. One such concept is the pervasive manifestation of self-fulfilling prophecies. Most individuals perceive their behavior to be volitional and independent of others’ expectations (see Wegner & Bargh, 1998). To address this challenge, an interactive learning task is an ideal method by which to help students appreciate the power of self-fulfilling prophecies.

Classic investigations of the self-fulfilling prophecy illustrate its potency (see Jussim, 1986). Merton (1948) proposed that self-fulfilling prophecies can occur when an individual behaves in such a way that his or her expectations of another individual are confirmed. In one of the earliest studies, Rosenthal and Jacobson (1966) informed a sample of school teachers that the performance of a particular group of students would improve sharply. The group of children (chosen randomly) did show significantly improved performance, leading the researchers to conclude that the teachers’ expectations elicited the students’ improvement.

Research on the self-fulfilling prophecy has extended beyond its implications for education. For example, Kelley and Stahelski (1970) found that people who believe others to be competitive provoke uncooperative behavior. In an early study on the behavioral confirmation of stereotypes, White participants interviewed Black and White “applicants” (Word, Zanna, & Cooper, 1974). Although the applicants’ responses were standardized, their performance was undermined by the interviewers’ behaviors. Investigators from another classic study (i.e., Snyder, Tanke, & Berscheid, 1977) tape-recorded male “perceivers” conversing with female “targets” who they believed to be either attractive or unattractive. Female targets who were labeled attractive behaved in a manner that was friendlier, more sociable, and more likeable than those targets labeled unattractive. Similarly, in a study on the stereotypes of obese individuals, condition-blind coders rated the behaviors of women in conversations with male partners who believed the women to be either thin or obese (Snyder & Haugen, 1995). Men’s expectations of their conversation partner’s weight and the associated stereotypes affected women’s behaviors. These studies elucidate the problematic role of self-fulfilling prophecies in a destructive cycle of stereotyping.

Students can extract a cursory understanding of self-fulfilling prophecies from a psychology textbook definition. An oral presentation of the aforementioned research could also enrich comprehension. However, an activity that allows students to observe a self-fulfilling prophecy as it unfolds is likely to be a much more powerful teaching tool (e.g., Benjamin, 1991).

Method

Participants

Eighty-one undergraduates (45 men, 36 women) from an introductory psychology course observed the classroom demonstration. After warning students that the task might be somewhat embarrassing, the instructor solicited 5 volunteers (3 men, 2 women) from the class to participate in the activity. An additional 20 students from a separate introductory psychology class served as a control group.

Materials

This demonstration required five standard baseball hats, labeled with dark ink and large letters: “good leader,” “very attractive,” “funny,” “annoying,” and “lazy.”

Procedure

To guide students’ attention during the activity, the instructor presented a brief description of the self-fulfilling prophecy to the entire class before requesting volunteers.
This presentation included a definition of the concept and a discussion of the historical context of research on the self-fulfilling prophecy and the relevant findings. The demonstration commenced when the instructor asked that volunteers come to the front of the classroom.

Without their knowledge of its content, a labeled hat was placed on each volunteer’s head. The instructor told the volunteers to treat each other in accordance with the information on their hat while working together on a series of group tasks. These tasks required volunteers to (a) name a new school mascot for their university, (b) determine the three best reasons for being a psychology major, (c) decide the distance between two major constructions on the school’s campus, and (d) line up in order of their presumed likeability as a function of the previous interactions. After each task, the instructor asked for the group’s answer and noted which volunteer responded with the answer before presenting the next task. After they completed the tasks, each volunteer guessed whether his or her label was positive or negative, what it said precisely, and what clues he or she used to make guesses. Finally, the instructor facilitated a class discussion to elicit specific examples of the self-fulfilling prophecy from the demonstration.

Results

Preliminary Evidence

As evidence of the communication of volunteers’ expectations for each other, each volunteer guessed the label of his or her hat. All five responded correctly regarding the positive or negative nature of their label, and all five correctly guessed the label itself.

Eliciting the Prophecy

Four assessment techniques indicated the degree to which the demonstration illustrated the self-fulfilling prophecy: (a) the observable behavioral results from the tasks served as indicators of the demonstration’s success, (b) the observers rated the degree to which the self-fulfilling prophecy emerged on the tasks with a Likert-type scale ranging from 1 (not a lot) to 5 (a lot), (c) after the demonstration volunteers were asked to rate the degree to which each volunteer behaved in accordance with his or her label, and (d) the observers recalled an example of the self-fulfilling prophecy from the demonstration. The results of the tasks provided evidence of the emergence of the self-fulfilling prophecy in the demonstration. When asked to decide on a mascot, after strained discussion, the “good leader” offered that she should be the new mascot. The “good leader” also reported that the results of the second task were that the top three reasons to be a psychology major were getting to wear hats, sleeping, and friends. During the third task, the volunteer wearing the “funny” hat began to make jokes about the volunteer wearing the “lazy” hat. For example, when determining the distance between two points, the “funny” volunteer asked the “lazy” volunteer, “Have you ever walked that far?” Volunteers lined up in order of likeability (attractive, good leader, funny, lazy, annoying) very quickly, without resistance from those wearing the negative hats. In fact, the “annoying” volunteer said loudly, “I guess I’m at the back.” Across tasks, the “good leader” consistently reported the group’s decision.

Overall, the observers rated the tasks to be effective in eliciting the self-fulfilling prophecy ($M = 3.93, SD = .72$). The observers also reported that the volunteers acted in accordance with their labels overall ($M = 3.78, SD = .84$). However, a repeated measures ANOVA suggested that some roles were more effective in evoking the self-fulfilling prophecy, $F(4, 77) = 17.24, p < .01$. The role that received the highest rating was “good leader” ($M = 4.00, SD = .85$), whereas the role that received the lowest ratings was “attractive” ($M = 3.01, SD = 1.02$).

The observers recalled an example of the self-fulfilling prophecy from the demonstration to qualitatively illustrate the activity’s effectiveness. The respondents recalled specific instances of the self-fulfilling prophecy in the demonstration. For example, one student wrote that “the good leader was naturally looked up to and took that role by providing good firm decisions that others could follow. Because others looked up to the leader she stepped into the role of a leader.” Another student responded, “The person wearing the funny hat would make wisecracks about the people around him, so in a sense people treating him as funny may have inspired confidence in him to actually be funny.” Yet another student wrote that “the annoying person seemed kind of annoying as a result of how people treated him.”

Overall Evaluation

We obtained pre- and postdemonstration ratings of the students’ knowledge of the self-fulfilling prophecy. To measure this knowledge, we averaged three knowledge items at each time with a response scale ranging from 1 (not a lot) to 5 (a lot). A sample item from this scale was, “How much do you know about the psychological phenomenon of the self-fulfilling prophecy?” A paired sample t test, $t(81) = −19.64, p < .01$, for this scale revealed that knowledge was significantly higher after the demonstration ($M = 3.89, SD = .76$) than prior to it ($M = 1.74, SD = .81$). To test students’ memory of the topic over time, the instructor included a multiple-choice question regarding the demonstration on the students’ first exam 2 months after its presentation. All of the 81 students (100%) who took the test answered the related question correctly.

An additional 20 students from two sections of a course in social psychology who did not do the learning activity also responded to this question. This class had covered the topic of the self-fulfilling prophecy in lecture and thus could serve as a control group for comparison to the experimental group. Almost a third of these students (30%) responded incorrectly when asked to choose the definition of a self-fulfilling prophecy. Students who witnessed the learning activity were significantly more likely to answer correctly than were students who were in the other classes, $χ^2(1, N = 101) = 25.84, p < .01$. Students who participated in the activity also reported that the demonstration taught them to think critically about
social interactions (M = 3.83, SD = 1.09) and that they found it to be interesting (M = 4.43, SD = .84), educational (M = 4.22, SD = .89), and enjoyable (M = 4.49, SD = .76). Finally, participants reported that they would recommend the activity to future classes (M = 4.40, SD = .85).

Discussion

The results of this study suggest that the concept of the self-fulfilling prophecy can be presented to psychology students effectively and enjoyably through an interactive learning task. The expectations assigned to volunteers by labeled hats elicited their behavioral confirmation. Both volunteers and observers provided evidence of the self-fulfilling prophecy’s emergence in the demonstration. The participants reported that the demonstration was interesting and educational. Finally, students’ knowledge of the self-fulfilling prophecy was significantly higher after the demonstration than before it.

Many potential applications exist for this demonstration. Depending on the research presented, the activity could be used in classes of introductory, social, industrial/organizational, educational psychology, or research methods. An introductory-level presentation should include reference to the experimental investigations in each area (see Jussim, 1986), whereas social psychology classes might focus more heavily on discussion of the self-fulfilling nature of stereotypes (e.g., Snyder et al., 1977). For the purpose of educational psychology, attention should be paid to Rosenthal and Jacobson’s (1966) original findings of the “Pygmalion in the classroom” effect and more recent replications (e.g., Madon, Jussim, & Eccles, 1997). The self-fulfilling prophecy as it applies to experimenter bias (e.g., Rosenthal, Persinger, Kline, & Mulry, 1963) is also relevant for discussion in research methods classes. Students in organizational psychology classes could benefit from the task as it relates to self-fulfilling prophecies in the workplace (e.g., Davidson & Eden, 2000).

One potential limitation to the study is that we compared responses to the multiple-choice item from the participants who took part in the demonstration to a convenience sample of students from a class that differed in its size and teacher sample. Hence, the classes may not be maximally comparable. Another potential limitation of this research is that all students who participated in the demonstration were aware of its goals. This knowledge may have induced demand characteristics on the part of observers and actors. However, although detracting from the experiment’s elegance, awareness of the demonstration’s goals enhances understanding by guiding students’ focus during the demonstration. Finally, students who participated in the demonstration may not have taken the tasks seriously. This lack of seriousness may arise, in part, from the experience of being in front of a large class and may be avoided by separating the class into smaller groups.

This article described an interactive learning task illustrating the self-fulfilling prophecy that has numerous classroom applications and provided strong empirical support for the demonstration’s effectiveness. The results of this study showed that students both learned from and enjoyed their participation in this task. Therefore, this demonstration provides a powerful tool with which teachers of multiple areas of psychology can enhance students’ understanding of the self-fulfilling prophecy.

References


Notes

1. A previous version of this article was presented at the 2001 Southwestern Conference on Teaching Psychology in Houston, TX.

2. We thank Sarah Burnet, Jennifer Knight, and Clare Reilly for their indispensable help in conducting this study.

3. Send correspondence to Michelle R. Hebl, Rice University, Department of Psychology, 6100 Main Street, Houston, TX 77005; e-mail: hebl@rice.edu.
Using *The Simpsons* to Teach Social Psychology

Judy Eaton and Ayse K. Uskul
York University

We examined students’ perceptions of the effectiveness of clips from the popular animated television show *The Simpsons* in illustrating key concepts in social psychology. Students rated the clips favorably and reported that the clips helped them understand the material better and apply social psychological concepts to real-life situations. In addition, students’ exam performance was significantly better on clip-related questions than nonclip-related questions. These findings suggest that television clips can facilitate the learning process.

Many instructors have found that showing all or part of popular films during class can increase student learning, interest, and enjoyment of key concepts by helping them make the connection between abstract theories and real-world examples (e.g., Badura, 2002; Boyatzis, 1994; Gee & Dyck, 1998; Kirsh, 1998; Raingruber, 2003; Roskos-Ewoldsen & Roskos-Ewoldsen, 2001). One disadvantage of using feature-length films is that they take up a significant amount of class time (Roskos-Ewoldsen & Roskos-Ewoldsen, 2001). An alternative to showing feature-length films is to use parts of a single television series to illustrate various key concepts throughout the course. In our undergraduate social psychology course, we showed clips from the animated television series *The Simpsons* to illustrate key social psychological concepts. We chose this particular cartoon for several reasons. First, we predicted that many students would be familiar with the show, which has been on television since 1989 and is also in syndication. Even if they did not watch the show, it was likely that students would be familiar with the characters and premise. We hoped this familiarity would decrease the set-up time for individual clips. Second, we expected that students had not thought about this particular show in an academic or critical way before. By examining social psychological concepts in novel ways, we hoped to increase students’ learning (Kirsh, 1998; Mathis & Tanner, 1991). Third, the cartoon provides a humorous look at various social situations. We hoped that the clips would make students laugh and have fun while helping them see the concepts depicted in more-or-less realistic situations. Research has shown that students respond to cartoon humor in a generally positive way (Lowis, 2002). In addition, the research on mood and learning suggests that positive moods are positively associated with certain kinds of learning (e.g., Ashby, Isen, & Turken, 1999; Fiedler, Nickel, Asbeck, & Pagel, 2003).

After identifying clips from the second season of *The Simpsons* (Groening, 2002; available on DVD) that could effectively illustrate key social psychological concepts (a complete list of which is available from the authors), we selected five of the most appropriate to show in class: one general clip depicting many different possible social psychological phenomena to present on the first day of class to generate discussion and four depicting specific concepts to present throughout the course (see Table 1). The length of each clip ranged from approximately 4 to 7 min. We were careful not to have clips from *The Simpsons* every class, partly because we did not want to overuse the technique and partly because we wanted to show other films and film clips.

During the first lecture, we told students that they were to watch a short video clip and their task was to identify any possible social psychological phenomena in the clip.

Table 1. Episode Clips from Season Two of *The Simpsons* and Social Psychological Concept Portrayed

<table>
<thead>
<tr>
<th>Episode Title and No.</th>
<th>Scene No., Title, and Description</th>
<th>Social Psychological Concept</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brush With Greatness (7F18)</td>
<td>No. 1: Main Title and No. 2: No One Gains 30 Pounds of Bone! A television commercial prompts Bart and Lisa to persuade Homer to take them to a water-themed amusement park. Homer gets stuck in a tube because he is too fat, is publicly humiliated, and vows to lose weight.</td>
<td>Persuasion, the self, prosocial behavior, social influence</td>
<td>5 min 46 sec</td>
</tr>
<tr>
<td>Bart vs. Thanksgiving (7F07)</td>
<td>No. 5: Now We Can Blame Him for Everything! Lisa tries to figure out why Bart ruined her Thanksgiving centerpiece.</td>
<td>Attributions</td>
<td>3 min 54 sec</td>
</tr>
<tr>
<td>Dead Putting Society (7F08)</td>
<td>No. 2: Marge, Beer Me! Homer is upset that his neighbor, Ned Flanders, seems to have a better life than he does.</td>
<td>Social comparison theory</td>
<td>3 min 50 sec</td>
</tr>
<tr>
<td>Three Men and a Comic Book (7F21)</td>
<td>No. 4: It Smells Like My Grandpa, and No. 5: If You Guys Hadn’t Tied Me, I Could Be Saving the Comic. Bart and two friends pool their money to buy a collectible comic and then fight over who gets to take it home.</td>
<td>Conflict resolution</td>
<td>7 min 20 sec</td>
</tr>
<tr>
<td>Itchy &amp; Scratchy &amp; Marge (7F09)</td>
<td>No. 2: I Told You, My Baby Beat Me Up. and No. 3: Dear Purveyors of Senseless Violence. After baby Maggie responds aggressively after watching a violent television program, Marge campaigns against the makers of the program.</td>
<td>Television violence and aggression</td>
<td>5 min 28 sec</td>
</tr>
</tbody>
</table>

*aScene numbers and titles correspond with those on the DVD (Groening, 2002). bClips start at the beginning of the scene.*

Vol. 31, No. 4, 2004 277
They then viewed the first *Simpsons* clip. The purpose of this exercise was twofold. First, research has shown that film clips shown on the first day of class can lighten the mood of the class and generate interest in the course (Badura, 2002). Second, we hoped that the clip would encourage class discussion. This exercise indeed seemed to lighten the mood of the class, and it was successful at encouraging students to participate in a class discussion.

We showed the remaining four clips throughout the course, immediately before introducing the relevant concept, a strategy recommended by Roskos-Ewoldsen and Roskos-Ewoldsen (2001). Following a brief discussion about what was depicted in the clip, the lecturer described the concept in detail, referring back to the clip when possible.

At the end of the course, students completed a questionnaire assessing their opinions of the various teaching aids used, including *The Simpsons* clips. We told students that their feedback would help the instructor assess and improve the effectiveness of various pedagogical aids, including the course Web site, the textbook, and other films shown in the course.

Students (*N* = 71) rated their agreement on a scale from 1 (strongly disagree) to 5 (strongly agree) with four items regarding the clips from *The Simpsons*. The items included “The use of the *Simpsons* clips was an effective way to illustrate key points” (*M* = 4.68, *SD* = .60), “The *Simpsons* clips helped me to understand the material better” (*M* = 4.41, *SD* = .86), “I enjoyed the use of the *Simpsons* clips” (*M* = 4.77, *SD* = .54), and “The *Simpsons* clips helped me to apply key concepts to real-life situations” (*M* = 4.11, *SD* = .85). These ratings suggest that students found the clips from *The Simpsons* to be effective at illustrating key points and in helping them understand the material better and apply the key concepts to real-life situations. Students also regarded the clips as highly enjoyable, as evidenced by both their high ratings and their written comments on the course evaluation. In addition, many students spontaneously mentioned how much they had enjoyed the use of the clips from *The Simpsons* in conversations with the instructor.

The fact that students enjoyed the clips and indicated that the clips helped them understand the concepts better is encouraging; however, student ratings do not allow us to determine whether the clips actually had a positive effect on student achievement. To address this issue, we examined students’ performance on multiple-choice exam questions related to the *Simpsons* clips. The midterm and final exams contained a total of 8 questions on the topics illustrated by the clips, and 142 questions on topics not illustrated by clips. We converted both scores to percentages for each student, and compared the mean performance. A total of 104 students wrote both exams. A paired-samples t test indicated that the percentage of correct answers on the questions relating to the clips (*M* = 83.17, *SD* = 13.58) was significantly higher than the percentage of correct answers on the questions not relating to the clips (*M* = 75.66, *SD* = 11.25), t(103) = 6.03, *p < .05*.

The combination of self-report and actual performance measures in this study provides converging evidence that the clips were effective at both generating student interest and increasing comprehension of the material. These results should be interpreted with some caution, however, as we were unable to control all potentially confounding factors in this classroom study. A fully experimental design would be a more powerful test of the clips’ effectiveness. Further investigation might involve comparing the *Simpsons* clips to other non-*Simpsons* clips, controlling for the amount of time spent on clip and nonclip topics in class, and testing other ways to use the clips (e.g., explaining the concept before showing the clip). Nonetheless, we believe that our results provide preliminary evidence that using television clips in the classroom can help the learning process. As suggested by an astute student in the class, one reason for the enduring success of *The Simpsons* may be its unique ability to tap into key social psychological concepts. We believe that this feature also makes it a potentially useful pedagogical tool for teaching social psychology.

References


Notes

1. We thank Randolph Smith and three anonymous reviewers for their insightful and helpful comments on a previous version of this article.

2. Send correspondence to Judy Eaton, Department of Psychology, Behavioural Sciences Building, York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3; e-mail: jeaton@yorku.ca.
Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.
What’s in a Name? Better Letters
If It’s Mine!

Angela Lipsitz
Lance A. Gifford
Northern Kentucky University

We describe a simple classroom demonstration of the name letter effect, the preference for letters in one’s own name (Nuttin, 1985), and provide support for the demonstration’s effectiveness.

Although popular psychology literature suggests that low self-esteem is a widespread problem, research literature indicates that, quite the contrary, most individuals think rather highly of themselves. In fact, people employ a vast arsenal of self-serving cognitions to maintain self-esteem at elevated levels. People rate their traits and abilities as above average (Alicke, 1985) and see the traits and abilities they possess as the ones most important to have (Dunning, Perie, & Story, 1991). People see their successes as caused by these traits and abilities and attribute their failures to other individuals, bad luck, or events beyond their control (Bradley, 1978). In addition, people exaggerate their roles in past events (M. Ross & Sicoily, 1979), overestimate their similarity to others (L. Ross, Greene, & House, 1979), and are unrealistically optimistic about future events (Weinstein, 1980).

So high is self-esteem that people even think the letters in their names are better letters. Nuttin (1985) called this finding the name letter effect and regarded it as a specific case of the mere ownership effect, the idea of valuing objects that are part of oneself more than objects that are not (Heider, 1958; see also, Beggan, 1992). Nuttin (1985, 1987) found that, if asked to choose their favorite letter from two- or three-letter groups or to choose their six favorite letters from a larger group, research participants chose letters in their first and last names at a higher-than-chance frequency. The effect has been demonstrated in over a dozen languages, including those using non-Roman alphabets (Hoorens, Nuttin, Herman, & Pavakanun, 1990; Nuttin, 1987). Researchers have shown that the effect is not due to name letters being more frequent (Nuttin, 1987), to an attachment to letters first written (Hoorens & Todorova, 1988), or to participants guessing the purpose of the research (Nuttin, 1985).

Like many of the other self-serving tendencies, the name letter effect can be demonstrated in class. Bourtight-Horowitz (1995) offered one method, but it is cumbersome to carry out, as it necessitates dividing the class into pairs, having half the class leave the room and having students construct test materials. We offer an alternative method, one that takes less time and is usually sufficient to demonstrate the effect.

Instructions

Conducting the demonstration requires about 20 min of class time and copies of a handout containing the letters of

Notes

1. Christopher Hakala is now at Western New England College.
2. Portions of this article were presented at the 69th annual convention of the Eastern Psychological Association, Boston, March 1998, and the 39th annual convention of the New England Psychological Association, West Hartford, CT, October 1999.
3. We thank Randolph Smith and three anonymous reviewers for comments that greatly improved the article.
4. Send correspondence to Howard C. Berthold, Department of Psychology, Lycoming College, Williamsport, PA 17701–5192; e-mail: berthold@lycoming.edu.
the alphabet listed vertically in random order. Beside each letter is, first, a rating scale, ranging from 1 (do not like) to 5 (like a lot), and then four blank lines under the column headings IYFN, NIYFN, IYLN, and NIYLN. Acknowledging that the task may seem rather silly, we ask students to rate rapidly how much they like each letter. Once all students have finished, we ask them to print their first and last names at the top of the handout. Then we explain the letters above the columns: IYFN stands for In Your First Name; NIYFN, for Not In Your First Name; IYLN, for In Your Last Name; and NIYLN, for Not In Your Last Name. Next, we ask the class to fill in their letter ratings under the appropriate columns. For example, if a student’s first name was Kevin, he would fill in his ratings for K, E, V, I, and N in the IYFN column. He would fill in his ratings for all other letters in the NIYFN column. If this student’s last name was Arnold, he would fill in his ratings for A, R, N, O, L, and D in the IYLN column and his ratings for all other letters in the NIYLN column. When all students have filled ratings in the appropriate columns, we ask them to calculate their mean for each column. After they complete this task, we ask the following question: “How many of you had a higher average for letters in your first name than for letters not in your first name?” The majority of hands usually go up. We repeat the question for last name with, typically, the same result.

At this point, we explain the name letter effect and mere ownership effect, plus the supportive research noted previously. Instructors tying the topic into a discussion of other self-serving biases could conduct demonstrations of those on the same day. Alternatively, instructors could use the name letter effect as an example of processing without conscious awareness.

**Evaluation**

We have found this technique to be successful on multiple occasions. Twice after conducting the demonstration in introductory psychology, we collected and double-checked students’ handouts and then calculated group means. In one class, in which 45 students participated, mean liking for letters in students’ first names was 3.7, and mean liking for letters not in their first names was 3.0. Mean liking for letters in their last names was 3.5 and for letters not in their last names, 3.0. In the other class, means for letters in and not in students’ first names were 3.8 and 3.0, respectively, and means for letters in and not in their last names were 3.6 and 3.0, respectively. Within-group t tests showed that all differences were significant at the p < .0001 level. In the first class, we also found that those present for the demonstration were more likely to answer a multiple-choice test question on the name letter effect correctly, \( \chi^2(1, N = 80) = 35.7, p < .0001 \), than those who were absent. Too few students were absent on the day of the demonstration in the second class to conduct a similar analysis although, of those present, over 80% successfully answered the test question.

In our experience, this is a simple and reliable demonstration, one that is useful for social, cognitive as well as introductory courses. It is also an interesting demonstration, as students are amused to learn of their latent letter preferences.

**References**


**Note**

Send correspondence and requests for copies of the handout to Angela Lipsitz, Department of Psychology, Northern Kentucky University, Highland Heights, KY 41099; e-mail: lipsitz@nku.edu.

**“Me Conform? No Way”: Classroom Demonstrations for Sensitizing Students to Their Conformity**

C. R. Snyder  
*University of Kansas, Lawrence*

This article presents 9 class demonstrations for increasing student awareness of personal conformity. One demonstration uses the Milgram (1965) obedience studies, a second and third involve self-fulfilling prophecies, a fourth and fifth illustrate the role of clothing, a sixth and seventh focus on peer pressure, an eighth examines individual differences in conformity, and a ninth provides a humorous example of conformity.
Why Does the “Above Average Effect” Exist? Demonstrating Idiosyncratic Trait Definition

Jason A. Nier  
Connecticut College

This article describes a classroom demonstration that illustrates idiosyncratic trait definition, a psychological mechanism that is partially responsible for the “above average effect.” For this demonstration, students rate themselves on a set of traits that vary in terms of ambiguity. The resulting self-evaluations should be more positive for highly ambiguous characteristics because these characteristics give individuals more leeway to construct a particularly favorable self-evaluation. The demonstration requires approximately 10 min to complete, and the instructor can discuss the results during the next class meeting as an illustration of the psychological processes responsible for the self-serving bias.

The above average effect is the tendency for individuals to see themselves as above average on characteristics that are both ambiguous and socially desirable (Myers, 2002). Although the above average effect has been documented for a variety of different characteristics and classroom demonstrations of the effect already exist (Bolt, 1998), there is no published procedure that allows instructors to illustrate the psychological processes responsible for the above average effect. This demonstration is a modification of existing classroom exercises that more clearly illustrates how individuals come to view themselves in such a favorable light.

Dunning, Meyerowitz, and Holzberg (1989) hypothesized that individuals construct idiosyncratic definitions of traits that allow for the maintenance or enhancement of a positive self-image. Specifically, when defining traits, individuals select criteria that define the traits in a manner that makes them appear especially favorable and disregard the criteria that make them seem less favorable. For example, when asked to evaluate themselves in terms of intelligence, individuals may focus on their exemplary grade point average but discount their poor performance on the SAT so that they may regard themselves as particularly intelligent.

An instructor can demonstrate the process of idiosyncratic trait definition by having students rate themselves on a set of characteristics that vary in terms of ambiguity. Highly ambiguous traits, by definition, are characterized by a wide range of behaviors. For these characteristics, individuals can construct an idiosyncratic trait definition with relative ease. Conversely, less ambiguous traits are defined by a narrower range of behaviors, and it becomes more difficult to construct idiosyncratic definitions of these traits. Therefore, if the idiosyncratic definition of traits is responsible for the above average effect, the effect should be greater for highly ambiguous traits because a wider range of behaviors can be called on to construct a more favorable trait definition.

Method

Eighty students completed a questionnaire that asked the following question: Compared to other college students of the same class level and sex as yourself, how would you rate yourself on the following characteristics? Students then rated themselves on a set of the 20 characteristics (see Table 1). As established by Dunning et al. (1989), these characteristics vary systematically in terms of their social desirability (positive traits and negative traits) and their ambiguity (high ambiguity and low ambiguity). Students made self-ratings on the following 9-point Likert scale ranging from 1 (considerably well below average), 2 (well below average), 3 (below average), 4 (slightly below average), 5 (average), 6 (slightly above average), 7 (above average), 8 (well above average), to 9 (considerably well above average). This scale is based on a demonstration by Bolt, 1998. The self-rating scale required approximately 10 min to distribute, complete, and collect.

During the next class meeting, I presented the results of the demonstration to the class as part of a class discussion of the self-serving bias. Following this discussion, I asked whether the demonstration contributed to their understanding of the origins of the above average effect and to what extent they thought that the demonstration should be repeated in future classes. Students indicated their responses on a scale ranging from 1 to 7, where higher numbers corresponded to a more favorable evaluation of the demonstration.

Results

As expected, there was a significant Social Desirability × Ambiguity interaction, F(1, 79) = 22.2, p < .001, which indicated that the above average effect was significantly larger for the highly ambiguous traits than for less ambiguous traits. Specifically, among positive traits, the degree to which participants saw themselves as above average (where an average self-evaluation was equal to 5) was larger for highly ambiguous characteristics (M = 6.1) than for less ambiguous characteristics (M = 5.6), t(79) = 4.6, p < .001. Similarly, among
negative traits, the degree to which participants perceived themselves to be below average was larger for highly ambiguous traits (M = 4.3) than for less ambiguous traits (M = 4.6), t(79) = 2.7, p < .01.

Students’ self-reported attitudes toward the demonstration suggested that the exercise was a worthwhile use of class time. Students indicated, on average, that the demonstration did indeed aid their understanding of the origins of the above average effect (M = 5.9, SD = 1.0). Also, when asked to indicate whether the demonstration should be repeated in future classes, students agreed that the demonstration should be repeated (M = 5.6, SD = 1.0).

Discussion

This exercise provides a procedure for demonstrating a psychological process that is partly responsible for the above average effect. Although this demonstration relies solely on self-report ratings, the process of idiosyncratic trait definition has fairly broad implications for social perception that reach well beyond self-rating scales. To drive home this point in the classroom, instructors may find it useful to explain how other potentially important forms of the self-serving bias may operate through idiosyncratic trait definition. For example, instructors can discuss whether the group-serving bias in intergroup perception (Pettigrew, 1979) may be partially due to idiosyncratic trait definition. When people evaluate the characteristics of the groups to which they belong, do they construct trait definitions that make their ingroups appear particularly favorable?

On a lighter note, a discussion of humorous examples of the above average effect can also be helpful in generating a lively discussion of the self-serving bias and idiosyncratic trait definition. For instance, the above average effect appears to extend to people’s perceptions of their pets. A survey of 1,179 pet owners conducted by the American Animal Hospital Association (Matheny & Miller, 2000) revealed what might be called a pet-serving bias. When asked to rate how intelligent their pet was, nearly all pet owners surveyed (99%) believed that their pets were at least average in intelligence. The genius rating was given by 18% of pet owners, 57% called their pets smart, 24% average, and just 1% below average. The existence of this above average effect can easily be explained in terms of idiosyncratic trait definition because the evaluation of a pet’s level of intelligence is quite ambiguous.

The process of idiosyncratic trait definition may extend not only to pets, but even into the afterlife, as a U.S. News & World Report survey found (Stanglin, 1997). When 1,000 Americans were asked who was likely to go to heaven (including themselves), 87% believed that they were likely to pass through the pearly gates, a percentage higher than any other individual that respondents evaluated. A distant second was Mother Teresa (79%), followed by Oprah Winfrey (66%), and Michael Jordan (65%). Could such a profoundly important self-perception be explained in terms of idiosyncratic trait definition? Perhaps individuals define their worthiness for the hereafter based on the criteria that assure their entrance into the afterlife and disregard their behaviors and beliefs that might otherwise damage their ability to gain access to heaven. Thus the effects of the idiosyncratic trait definition may have implications for even the most fundamental and enduring metaphysical questions about ourselves.

References


Note

Send correspondence to Jason A. Nier, Box 5305, Department of Psychology, Connecticut College, 270 Mohegan Avenue, New London, CT 06320; e-mail: janie@conncol.edu.

Two Active Learning Exercises for a History of Psychology Class

David Zehr

Plymouth State University

History of psychology students participated in 2 active learning exercises. The first exercise focused on William James. The students role-played members of a psychology department faced with the question of hiring James for an open position in the department. The second exercise, used in teaching the history of applied psychology, combined role playing and a variation of the currently popular 7-min speed-dating technique. Students found both exercises interesting and helpful in understanding course content and recommended using them in future classes.

Keeping a history of psychology course fresh, relevant, and engaging for students is important given that students often enter the course with limited knowledge (Brewer & Davis, 1999) and preconceived notions of irrelevance. Many teachers encourage the use of active learning exercises to overcome such obstacles (Henderson, 1995), and previous research suggests that such strategies can be effective at enhancing students’ understanding of and appreciation for historical issues (Carpenter, 1990; Howard, 1990; Zehr, 2000). Toward this end I developed two active learning exercises to present course material. The first exercise focused on William James and the second on the development of applied psychology in America. I used both exercises for three consecutive semesters in a history of psychology course.

---

54 Teaching of Psychology
Faculty Forum
Bryan K. Saville, William Buskist, Robert Weis, Jane P. Sheldon, Stephen J. Dollinger, Lisa Curtin, Denise M. Martz, Doris G. Bazzini, Barbara Bowers Vicente, Jason A. Nier, David Zehr, Carmen Garcia & Maria Teresa Ruiz Garcia

Available online: 20 Nov 2009

To cite this article: Bryan K. Saville, William Buskist, Robert Weis, Jane P. Sheldon, Stephen J. Dollinger, Lisa Curtin, Denise M. Martz, Doris G. Bazzini, Barbara Bowers Vicente, Jason A. Nier, David Zehr, Carmen Garcia & Maria Teresa Ruiz Garcia (2004): Faculty Forum, Teaching of Psychology, 31:1, 41-58

To link to this article: http://dx.doi.org/10.1207/s15328023top3101_10

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
I Scream, You Scream: Teaching Validity and Reliability Via the Ice Cream Personality Test

Marianne Miserandino
Arcadia University

This exercise uses the Internet-based “Ice Cream Personality Test” to help undergraduates understand the principles of personality testing including reliability, validity, Barnum statements, and generalizability. Results indicated that the Ice Cream Personality Test, although great fun, lacked reliability and validity. Students found this exercise enjoyable, useful, thought provoking, and apt to make them skeptical about personality tests they might encounter on the Internet or elsewhere.

Does teaching about validity and reliability make your students scream? The Internet is filled with all sorts of entertaining questionnaires purporting to be personality tests, but are these tests legitimate? The Ice Cream Personality Test (http://www.personalityquiznet/test/icecream.htm1), which originally appeared on the Edy’s ice cream home page (http://www.edys.com) and yielded more than 23,000 hits on Google (accessed on January 22, 2005), contains playful graphics and humorous feedback. These qualities make the test an entertaining and effective way for undergraduates to experience the importance of validity, reliability, and generalizability of personality tests. The Ice Cream Personality Test piques students’ interest and challenges them to think critically to discover what makes this false test so appealing.

The designers of the Ice Cream Personality Test claim that based on “psychological research” they can tell a person’s personality from that person’s favorite ice cream flavor. On the screen, respondents choose their favorite flavor by clicking on a cartoon of one of six ice cream cones. Once they make their choice, respondents receive a few lines of feedback describing their personality. The test is fun to take, and visitors report that their results ring true.

There are only a few published classroom exercises that demonstrate the importance of validity and reliability in personality testing. For example, to learn about testing, students might rate the personality of the professor (Reinehr, 1991), identify their astrological trait profile (Ward & Grasha, 1996), or design their own scales (Benjamin, 1987; Camac & Camac, 1993; Davidson, 1987; Hynan & Foster, 1997). Similarly, demonstrations have used the Barnum effect to illustrate the importance of specific feedback and research ethics (Baillargeon & Danis, 1984; Beins, 1993; Boyce & Geller, 2002; Forer, 1949; Furnham & Schofield, 1987; Holmes, Buchannan, Dungan, & Reed, 1986; Russo, 1981). Despite the ubiquity and popularity of demonstrations, there are none that encourage students to be skeptical of Internet personality tests.

Students in my personality psychology class administered the Ice Cream Personality Test to people they knew and tried to predict their friends’ favorite flavor and personality to judge the validity of the test. The compiled class’s results and their answers to the assignment became the basis of a class discussion on issues in personality testing.

1This Web site recently changed the graphical format of the test. See http://www.mda.state.mi.us/kids/countryfair/games/icecream/icecream.asp for the original ice cream cone graphics.
Method

After covering personality tests, reliability, and validity, I gave students in both sections of my personality psychology class an assignment in which they gave a pencil-and-paper version of the Ice Cream Personality test to a sample of five of their friends (see Appendix). Unbeknownst to the students, they presented the statements and flavors to their participants in a randomized order. Students attempted to guess their friends’ personality and favorite ice cream flavors, calculated their hit rate, and compiled the results in a table that was formatted so that students would be able to see clusters of hits along the diagonal, if there indeed was a relation between personality and flavor preference, as the test claims (see Table 1). Finally, after considering the pattern of their participants’ responses, their success or failure at guessing their participant’s responses, the relation between the results of this test and other standard personality tests studied earlier in the course (e.g., the Eysenck Personality Inventory [Eysenck & Eysenck, 1964], the Big Five trait inventory [Costa & McCrae, 1992]), students evaluated the test in a one- to two-page paper.

Results

Table 1 presents the results based on the class’s data. There are 180 data points from 36 students who each interviewed 5 participants. Because of the low expected frequencies for some of the cells, a chi-square test was not an appropriate way to test for a relation between flavor choice and personality description. However, a related statistic, Cramer’s Phi (\( \phi = .22 \)), indicated that flavor choice accounted for only 5% of the variance in personality descriptions (Cramér, 1946). Even by visually examining Table 1, there were no more hits along the diagonal nor any other pattern (except the overwhelming preference for chocolate ice cream), indicating that the personality test lacks validity.

Discussion

Reliability

Reliability is the extent to which a test consistently reflects the characteristic being studied (Wiggins, 1973). Students can see if the test has high test–retest reliability by reading the test to each participant after a short interval (e.g., 1 week). Because the participants’ personalities are the same, they should receive the same feedback about their personalities, even if they choose a different flavor. However, that is not the case because choosing a different flavor yields a completely different personality description. Therefore, the test lacks the reliability to measure personality accurately.

Validity

Validity is the extent to which a test measures what it is supposed to measure (Cronbach & Meehl, 1955). There are many different kinds of validity (see Jackson & Paunonen, 1980), but the ones most relevant to this exercise are construct validity, criterion validity, and convergent validity. If a test measures the intended theoretical concept, then the test has construct validity (Cronbach & Meehl, 1955). In the case of the Ice Cream Personality Test, there was no relation between the flavors and personality descriptions participants picked.

If a test predicts to a standard external to the test it has criterion validity. This criterion may be in the future (predictive validity) or at the same time as the test (concurrent validity; Cronbach & Meehl, 1955). Students were unable to correctly guess either a participant’s personality description or favorite flavor, indicating a lack of validity by these types of validity also.

If a test correlates with other related measures (Campbell & Fiske, 1959), then the test has convergent validity. Personality descriptions from the Ice Cream Personality Test should correlate with their scores on the Eysenck Personality Inventory or the Big Five trait inventory. Although students scored similarly on these latter two measures, their results were not correlated with results from the Ice Cream Personality Test.

Barnum statements. Barnum statements are descriptions or predictions that are so general or ambiguous that they could apply to almost anybody, thereby indicating a lack of validity (Dickson & Kelly, 1985; Forer, 1949). Often, people do not recognize such statements as generalities and are easily taken in by such feedback. Here, the personality descriptions purportedly go with a specific ice cream flavor are all positive in nature, can apply to many people, and are broad enough to be interpreted by the participant as accurate.

Generalizability

Generalizability is the extent to which a test maintains its validity across contexts, including different groups of people and different conditions or situations (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Wiggins, 1973). The Ice Cream Personality Test claims to be valid for all populations and conditions, yet the test is not applicable to people who have food allergies, are lactose intolerant, on diets, or who do not like ice cream. Although these conditions are obviously not

---

2This assignment was 1 of 10 homework assignments during the course of the semester. A percentage of the homeworks were optional but could earn students extra credit. Indeed, 36 students out of 52 students chose to do this homework assignment.

3These personality tests were readily available in the instructor’s manual (Shackelford, 2005) for our textbook (Larsen & Buss, 2005).
relevant to personality testing, using these examples demonstrates for students why generalizability is an important issue. Furthermore, although our results suggest the test lacks validity, given our small and nonrepresentative sample we should not generalize the results and claim that the test lacks validity for other samples.

Student Evaluation

On a scale ranging from 1 (not at all) to 5 (very), with 3 (somewhat) at the midpoint, a majority of students\(^4\) (N = 42) rated the homework assignment and class discussion above the midpoint on six evaluation questions. Specifically, they found it an enjoyable (M = 3.8, SD = .95) and useful way to learn about validity and reliability (M = 4.2, SD = .75). They reported that the activity was thought provoking (M = 3.9, SD = .72) and made them think about the importance of validity and reliability in test design (M = 4.3, SD = .66), and it even made them skeptical about personality tests they might see in magazines or on the Internet in the future (M = 4.5, SD = .71). One student remarked that “The Ice Cream Personality Test is nothing but a hoax!”

Overall, an overwhelming majority of the class (37 students) recommended using the Ice Cream Personality Test in future classes. One enthusiastic student said “This was a very fun way to get the concept.” Another student said that this assignment was particularly helpful to gain “a better understanding of the readings.”

The Ice Cream Personality Test provided students with a humorous and interesting way to understand and apply principles of psychological testing. Although the test falls short as a personality measure, it measures up to a fun way to learn about concepts that students otherwise find dull.

\(^4\)An additional six students were present in class on the day we discussed the assignment and evaluated the assignment even though they did not do the assignment.

---

**Table 1. Compiled Results of the Ice Cream Personality Test**

<table>
<thead>
<tr>
<th>Personality Description No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorite Flavor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vanilla</td>
<td>9</td>
<td>11</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>Chocolate</td>
<td>6</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>11</td>
<td>55</td>
</tr>
<tr>
<td>Butter pecan</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Banana</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Strawberry</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>Chocolate chip</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>11</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>29</td>
<td>24</td>
<td>37</td>
<td>26</td>
<td>38</td>
<td>180</td>
</tr>
</tbody>
</table>

*Note. Numbers along the diagonal indicate the predicted combination of flavor with personality description.*

---

**References**


Implementing an Undergraduate Laboratory Course in Functional Magnetic Resonance Imaging

Kevin D. Wilson
Gettysburg College

This article describes the feasibility of implementing a functional magnetic resonance imaging (fMRI) laboratory course at an undergraduate-focused institution without internal scanning facilities. I discuss how to incorporate specific functional brain imaging topics into lectures, how to design and implement laboratory sessions that allow students to analyze existing fMRI data sets, and how to incorporate empirical research projects involving novel fMRI data collection into the course through collaborations with researchers at larger institutions. This type of course is possible at virtually any institution and provides an excellent opportunity for advanced undergraduate students to gain first-hand research experience in cognitive neuroscience.

An important goal for liberal arts science education is to provide students with research opportunities that will position them competitively for admission to graduate school (Kardash, 2000; Kremmer & Bringle, 1990). Although laboratory courses are more prevalent at Research I universities than at undergraduate-focused institutions (Messer, Griggs, & Jackson, 2000), smaller schools often are able to provide comparable experiences in many areas of psychology. Disciplines such as cognitive neuroscience present a host of challenges to this prospect. Specifically, many cognitive neuroscientific techniques are expensive and require support systems and personnel that are often unavailable at undergraduate-focused institutions. Nevertheless, introducing students to these methodologies is critical for preparing the next generation of scientists.

Research involving functional brain imaging provides an excellent example of this challenge. Functional magnetic resonance imaging (fMRI) has exploded in the past decade (Cabeza & Nyberg, 2000; Fellows et al., 2005) and has become the dominant methodology in cognitive neuroscience. MRI scanners, however, are available only at Research I institutions and require dedicated personnel to operate (e.g., technicians, engineers, physicists). Furthermore, data collection can be prohibitively expensive ($600 or more per hr), and data analysis typically requires significant hardware and software investments that are beyond the budgetary limits of many smaller institutions.

Note
Send correspondence concerning this article to Marianne Miserandino, Department of Psychology, Arcadia University, 450 South Easton Road, Glenside, PA 19038; e-mail: miserandino@arcadia.edu.
Comedian Johnny Carson was a part of mainstream American TV and culture for over 30 years as host of the Tonight Show before Jay Leno. His life in the public eye and his recent death provide an opportunity to analyze the man. His The New York Times obituary provided a rich description of his personality and his life (Severo & Carter, 2005). Friends described Carson as “affable, accessible, charming and amusing” and yet off camera as “testy, defensive, preoccupied, withdrawn … inept and uncomfortable with people” (p. 1).

These descriptors are immediately recognizable as traits. Because obituaries describe a person and his or her life, they are rich in trait descriptions. In contrast, an obituary is not the time to discuss deeper psychoanalytic motivations or humanistic strivings. Many obituaries are mini case studies of regular, rather than abnormal, personalities.

Although there are many ways of categorizing human personality into traits, the Five Factor Model (FFM) has garnered tremendous support in recent years and features prominently in most introduction to psychology texts (e.g., Gerrig & Zimbardo, 2005; Gleitman, Fridlund, & Reisberg, 2004) and many personality texts (e.g., Larsen & Buss, 2005; Pervin, Cervone, & John, 2005). Although there are various conceptualizations, the model that often appears in undergraduate texts is by McCrae and Costa (1999). Their FFM is important because of its extensive use in varied research contexts, including personality of presidents (Rubenzer, Faschingbauer, & Ones, 2000), mate selection and marital satisfaction (Botwin, Buss, & Shackelford, 1997), likelihood of divorce (Kelly & Conley, 1987), music preference (Rentfrow & Gosling, 2003), and even perceptions of personality based on office space or apartment decoration (Gosling, Ko, Mannarelli, & Morris, 2002).

The five factors of personality are extraversion, agreeableness, conscientiousness, emotional stability, and openness-intellect, each made up of facets (Costa & McCrae, 1992). These facets and factors appear to be universals in that they can describe human personality across many times, samples, and cultures (McCrae & Costa, 1997; Yik & Bond, 1993).

Heeeere’s Johnny: A Case Study in the Five Factor Model of Personality

Marianne Miserandino
Arcadia University

I describe an assignment for personality psychology or introduction to psychology classes in which students used the Five Factor Model of personality to analyze the personality of entertainer Johnny Carson through his The New York Times obituary. Students evaluated this assignment highly: A majority indicated that the assignment was interesting, enjoyable, and useful in helping them to understand and apply the Five Factor Model, and all agreed that the assignment was thought-provoking.
(Mueller, 1985), students’ own behavior (Dunn, 1997; White, 1974), and popular music (Hughes, 1984) to illustrate principles of personality.

Only one of these methods, however, illustrated trait theories and behavior over time (Carlson, 1992). None discussed the FFM or personality coherence in particular. The majority did not even analyze an actual person. Using an obituary from a newspaper as reliable as the *The New York Times* (which is readily available online) minimizes problems such as privacy, reliability of the source, and the time required to view a movie or to read a book. Indeed, the assignment is short enough to take up a single class period.

Although students can resort to memorizing the FFM, this assignment gives them a chance to see how a person’s traits play out, change over time, and interact with other traits to form a coherent whole—and a very much real—personality. Because of the unusual number of trait terms used, Carson’s obituary is particularly apt for analysis with the FFM. Overall, this assignment using Johnny Carson’s obituary as a “case study” of the FFM serves all these ends and provides a shorter, more in-depth alternative to previously published assignments.

The Assignment

Students in both sections of a personality course (N = 52) received a copy of the obituary and the assignment.1 Students addressed a set of questions (see Table 1) in a 2- to 3-page paper. They used Carson’s actions, things he said, and things acquaintances said about him to evaluate his personality, quoting exact passages to support their view. The assignment also directed students to the official Johnny Carson Web site (http://www.johnnycarson.com/) and others.

On the day the papers were due, students used their individual papers as the basis of a class discussion, which functioned as a review before the exam on this material. Students divided into groups; each group discussed one of the following topics: neuroticism, extraversion, openness, agreeableness, conscientiousness, and personality coherence and change across Carson’s life. Six groups of three to five people is ideal; however, the instructor can easily condense or expand the topics to fit the number of students. Each group presented their opinions and evidence to the class, and through discussion we came to a consensus about Carson’s personality.

### Johnny Carson and the FFM

Table 2 presents facets of the FFM as illustrated by the personality of Johnny Carson. Overall, the class judged Carson as high in neuroticism, openness, and conscientiousness and low in extraversion and agreeableness.

Although Carson acted extraverted in public, evidence from the obituary suggests that he was really more of an introvert. For example, he “zealously guarded his private life” and was a “mystery man” to the public (Severo & Carter, 2005, p. 1). The difference between Carson’s public persona and private personality made extraversion particularly challenging for students to evaluate because it raised the question of who the real Carson was.

**Personality Stability**

Over the course of his lifetime Carson appeared to remain above average in humor, intelligence, and being a loner, showing personality stability in these traits.

**Personality Change**

There was not a lot of evidence for change in Carson’s traits, perhaps because an obituary writer strives for coherence. However, there was some evidence that Carson became more neurotic as an adult, experiencing increased anxiety and becoming more reclusive in his later years.

---

1 This assignment was one of three assignments during the course of the semester together worth 11% toward their final course grade. I required students to complete all three but dropped the lowest one.

---

**Table 1. Assignment Questions**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What can we say about Johnny Carson’s disposition?</td>
<td>Was he neurotic or emotionally stable? Extraverted or introverted?</td>
</tr>
<tr>
<td></td>
<td>Open or conventional? Agreeable or disagreeable? Conscientious or aimless?</td>
</tr>
<tr>
<td>In other words, based on the evidence presented in the article, where would a personality psychologist place Carson on each of the five dimensions of the FFM? On which facets would he be particularly high or low?</td>
<td></td>
</tr>
<tr>
<td>What does the life of Johnny Carson illustrate about the stability and change of human personality? What stayed the same and what changed about him over the course of his life?</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** FFM = Five Factor Model.
Personality Coherence

In studying a life, one can see the operation of personality coherence, where people may maintain their relative standing on a trait but manifest that trait differently at different times. When he was 12, Carson took to entertaining friends and family by doing magic tricks as “The Great Carsoni.” In college, he wrote his senior thesis on comedy writing. Soon after, he landed his first job in TV, which eventually led to his legendary post at the *Tonight Show* where he portrayed the magician “Carnac The Magnificent.”

Assessment of Student Learning

To measure the effectiveness of this assignment on students’ mastery of the material, I compared grades on the essay portion of the exam on this material. The essay questions from spring 2004 and spring 2005 were both case studies in which students analyzed a person using the FFM and discussed personality change, stability, and coherence. I graded essays out of 10 points using comparable grading criteria. Students in the section (*n* = 38) who analyzed Carson using the FFM scored higher (*M* = 7.4, *SD* = 2.11) on the essay question than students (*n* = 17) who did not have this experience (*M* = 5.91, *SD* = 2.24), *t*(53) = 2.40, *p* = .02, even though the exam grades were higher for the spring 2004 section. This finding provides evidence that the Johnny Carson assignment helped students master the FFM and related concepts.

Student Evaluation

Despite the fact that this assignment was due immediately before an exam, students were enthusiastic. On a scale ranging from 1 (not at all) to 5 (very), with somewhat at the midpoint (3), students found the assignment interesting (*M* = 3.7, *SD* = .78), enjoyable (*M* = 3.5, *SD* = .84), thought-provoking (*M* = 4.3, *SD* = .55), and a useful way to apply the FFM (*M* = 4.3, *SD* = .62). Indeed, in response to the question “Did this activity make you think?” 100% of the class agreed that it did, despite the fact that 3 to 6 of the 52 students consistently rated the activity below the midpoint on the interesting and enjoyable items.

Of the six students who wrote comments, five were positive. For example, one student said “It really helped me to apply and understand what I was learning.” Another summed up what active learning is all about: “It was a lot easier for me to compound the facets from the FFM by thinking in terms of the activity. Most likely, I would’ve just tried to memorize them all had it not been for this activity.” One student admitted “Did not really enjoy the paper … but the class [discussion] was useful.” The lone negative comment had more to do with the timing of the activity: “Add[ed] to the stress of the test. Should be [an] optional [assignment].”

<table>
<thead>
<tr>
<th>Table 2. The Five Factor Model Applied to Johnny Carson</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>Emotional Stability</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Openness</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Vol. 34, No. 1, 2007
Before this activity I thought that students would be excited by a chance to apply what they have learned to a celebrity. However, many students did not know who Johnny Carson was or never saw him. Instead of being problematic, their lack of familiarity with Carson actually made the assignment more useful, as all students came to the assignment with a lack of preconceived notions about Carson. One student remarked “For the most part, people didn’t know him so it was very effective.”

Applications

This assignment is readily adaptable to both introductory psychology classes and personality classes. Although I used it as a group assignment, one could readily make it into an individual assignment, an essay question, or even a full-class discussion. Furthermore, the students themselves might suggest deceased public figures whom they would like to study via an obituary.

References


Note

Send correspondence to Marianne Miserandino, Department of Psychology, Arcadia University, 450 South Easton Road, Glenside, PA 19038; e-mail: miserandino@arcadia.edu.
We describe the use of academic guide maps to learn psychology. Academic guide maps are the combination of derived structural schemas (recurrent knowledge structures in psychology) and node-link knowledge maps (diagrammatic presentations that illustrate the relational organization of information). We identified theories, research, procedures, techniques, classification systems, concepts, principles, processes, and systems as recurrent structures in psychology and created guide maps for each. Based on studies of derived structural schemas and node-link maps, the guide maps should help students when taking notes and during lectures and studying for exams. Additionally, the guide maps may aid professors in preparing and organizing their lectures and in assessing their students’ understanding of the material.

An analysis of introductory textbooks revealed that from 1988 to 1997 the size of introductory textbooks had grown by approximately 15%, from an average of 585 to 674 pages of text and 33 to 39 pages of text per chapter (Griggs, Jackson, Christopher, & Marek, 1999). To aid students in learning all of the information presented in textbooks, many texts now come with online and CD-ROM tutorials in addition to hard-copy study guides. Many of these learning aids use various types of drill and practice questions to help students assess their mastery of the material they have read. Of course, these drill and practice tools can be helpful, but only after students have attempted to encode the material. However, we have developed academic guide maps to aid students in both encoding and retrieving studied materials. The academic guide maps serve as advance organizers to guide and facilitate the assimilation of the materials being studied and provide a spatial organization that aids in encoding and retrieval of the material (for reviews, see Dansereau, 1995; Mayer, 1979; O’Donnell, Dansereau, & Hall, 2002).

Academic guide maps take advantage of the strengths and potential synergy of derived structural schemas and node-link knowledge maps. Schemas are well-developed mental representations of generalized knowledge in a particular domain (Rumelhart & Norman, 1978) that impose organization on newly encountered information and provide a preexisting framework for economically storing newly encountered information (e.g., Anderson, Spiro, & Anderson, 1978; Brooks & Dansereau, 1983; Dansereau, 1995; Rumelhart & Norman, 1978). People typically develop schemas via repeated experiences with similar objects and events (Rumelhart & Norman, 1978); however, experts can derive and label schemas to show the nonobvious, common elements and, therefore, reduce the number of experiences novices require to develop the schemas (for examples of derived schemas, see Brooks & Dansereau, 1983; Ohlhausen & Roller, 1986). In science education, for example, Brooks and Dansereau (1983) derived a scientific theory schema, DICEOX. DICEOX consisted of a description of the theory (D), the inventors and history (I), the consequences the theory has had on science and society (C), the evidence for and against the theory (E), other complementary or competing theories (O), and any extra information that did not fit under the D, I, C, E, and O categories (X). Brooks and Dansereau found that students who trained to use the DICEOX schema and who filled in the DICEOX categories as they studied passages on plate tectonics freely recalled more main ideas from the passages than students who used their own studying and test-taking methods (e.g., standard note-taking).

Thus, given the benefits of using DICEOX, we believed that students of psychology could also benefit from knowing a collection of structural schemas relevant to psychology. We, therefore, surveyed various introductory psychology textbooks, consulted with psychology instructors, and subsequently derived six schemas: theory, research, technique/procedure/classification system, concept/principle, process, and system.1 To present these structural schemas, we used node-link knowledge map formats that illustrate the relational organization of information. In node-link maps, subject matter placed in nodes is connected to related subject matter through links (e.g., see Figures 1 and 2). Research comparing using knowledge maps versus studying text shows that using maps leads to (a) better recall of main ideas, (b) greater self-reported motivation and concentration, and (c) greater learning benefits for students with low verbal skills.

1The following are the components of the six schemas: theory: description, history/contributors, evidence for and against, importance, similar/analogous theories, and competing theories; research: purpose, methodology, findings, implications, and alternative interpretations; technique/procedure/classification system: purpose, description, benefits, shortcomings, alternatives, and inventors; concept/principle: description, supporting research, importance, related concepts/principles, and related theories; process: purpose, description, potential problems, research, and similar processes; and system: function, parts, function of parts, potential problems, and supporting research.
To fully appreciate a scientific theory, you should be able to describe the theory, and you should also know the history of the theory, know evidence for and against the theory, know why the theory is important, and know whether there are any similar and competing theories.

We designed the academic guide maps to be flexible, modifiable, and general-purpose tools. We encourage the modification of our maps by adding, renaming, and dropping nodes and adding sketches to the maps when appropriate. We included a Quick Links node on each map to facilitate the integration of the information contained in the guide maps with the general topic, textbook readings, lectures, and personal experiences.

We believe that the use of academic guide maps should direct learners toward mastery goals: discrimination, monitoring, and integration versus repeated reading and rehearsing (Pintrich & Garcia, 1991). For example, we designed tutoring, and integration versus repeated reading and rehearsal.

Forgetting and Spawned Lines of Research

Ebbinghaus showed that longer retention intervals led to less savings. His research provided a method for systematically studying forgetting and spawned lines of research in memory as a mechanism for forgetting. Forgetting was systematic. Ebbinghaus was interested in all aspects of memory and forgetting systematically.

Who contributed to this theory?

Why is the theory important?

What is the purpose of the research?

What methodology was used?

What were the implications of the study?

Are there any alternative interpretations?

What was the purpose of the research?

Ebbinghaus was interested in all aspects of memory and forgetting systematically.

What methodology was used?

Who contributed to this theory?

Why is the theory important?

What is the purpose of the research?

Ebbinghaus served as his subject of study. Ebbinghaus learned lists of nonsense syllables (like ZAK) until he could produce the list twice without error. After a retention interval, he relearned the list. He measured how many trials (tests through the list) he took to relearn the list.

What were the implications of the study?

Ebbinghaus showed that longer retention intervals led to less savings or greater forgetting. His research provided a method for systematically studying forgetting and spawned lines of research on memory as a mechanism for forgetting.

Figure 1. Theory map. An example of a completed theory map on Piaget's theory of cognitive development.

Figure 2. Research map. An example of a completed research map on Ebbinghaus's studies of memory and forgetting. (for review, see O'Donnell et al., 2002). Thus, we developed academic guide maps for the six schemas that we derived for (examples of the theory and research maps, see Figures 1 and 2, respectively; see Note 1).

We designed the academic guide maps to be flexible, modifiable, general-purpose tools. We encourage the modification of our maps by adding, renaming, and dropping nodes and adding sketches to the maps when appropriate. We included a Quick Links node on each map to facilitate the integration of the information contained in the guide maps with the general topic, textbook readings, lectures, and personal experiences.

We believe that the use of academic guide maps should direct learners toward mastery goals: discrimination, monitoring, and integration versus repeated reading and rehearsing (Pintrich & Garcia, 1991). For example, we designed the questions within the nodes and the arrangement of the nodes on the research map (see Figure 2) to quickly engage learners in searching for relevant information while studying and taking notes. The questions within the nodes and arrangement of the nodes prompt the learner to look for the purpose of the study first, then for the methods and results, and then for the implications of the study and for alternative interpretations. To fill in each node, students must find information relevant to a particular study and discriminate the information relevant to each node. Students can then use the map to monitor the completeness of their understanding by assessing whether they have filled in all of the nodes. Finally, by completing the map, students should see how the information contained in each node contributes to a more complete understanding of the study; by completing a set of maps for the content area of the study, students should see how the particular study contributes to a more comprehensive understanding of the overall topic.

The use of academic guide maps should also promote more active participation in the classroom, particularly if professors use the maps to organize their lectures and as supplemental materials. When taking notes, students should see what information is missing from their maps and then ask appropriate questions. The use of maps, therefore, should empower the students to ask more questions about topics being presented and, thus, get students more involved in the learning process.

The use of maps similar in design to our guide maps has been effective in drug education and counseling (e.g., Dees, Dansereau, & Simpson, 1994), learning biology (Skaggs, Rewey, & Paulus, 1990), and aviation incident reporting (Wiegmann & von Thaden, 2003). Additionally, Cauchy and Dansereau (1996) and Jacobs-Lawson and Hershey (2002) suggested having students develop knowledge maps instead of or in addition to traditional essay assignments. Finally, one of the authors (Dansereau) has used more rudimentary versions.
of some of our academic guide maps in teaching techniques of college learning, general psychology, and memory and cognition courses, and students' feedback suggests that these types of guide maps are useful in organizing information and making connections between topic areas.

References


Notes


2. Send correspondence to Michael A. Motes, Department of Psychology, TCU Box 298920, Texas Christian University, Fort Worth, TX 76129; e-mail: m.a.motes@tcu.edu.
tudents to preview the chapter and engages them in creating the game. Furthermore, it gives the instructor time to make the cards. Second, from the students’ work I selected quality FWs. Third, students played the game in one class session prior to a test. An abnormal psychology class completed an anonymous evaluation of this activity.

Method

Timeline

I use FWs as review before exams (class size = 35 to 65). It does not replace primary instruction. About 2 weeks before an exam, I review a handout and assign four terms to each student, due at the next class. For each term, students generate five FWs and briefly explain why they chose them. In larger classes more than one student may generate FWs for a term; overlap provides the instructor variety when selecting quality FWs. Note that students may be generating FWs for terms not yet covered in lecture, but by the time the cards are ready for play, the students will be familiar with the terms (either through lecture or reading). Most students are able to glean information from the text to make adequate FWs.

Selecting FWs

To create quality FWs, students need to extract relevant information about the term from their text. For example, in creating the cards, it may be helpful for student and instructor to ask, “What are the essential elements related to acetylcholine?” and “What words would a player find most useful in describing it to another player?” Quality FWs are ones that most players would like to use and are integrally related to the term. For example, based on my text, students should note that acetylcholine is a neurotransmitter that plays a role in memory, muscle contraction, and Alzheimer’s disease. Thus, the FWs mentioned earlier are of high quality. FWs should include some understanding of important functions, description, purpose, or synonyms. After examining the student-generated words, I selected five exemplary FWs and made the cards. I did not limit the FWs to the use of single words, allowing phrases such as “father of psychology.”

For a class of 65 students I made cards for six groups, each comprised of two teams having 4 to 6 players on each team. For each group, I mixed and divided the cards, giving each team half. The students played the game, and a class evaluation (team half. The students played the game, and a class evalu-

ation resulted in FWs that were too concrete (for the term cones, FWs were rods, color, eye, retina, and daylight). Others placed too much emphasis on the textbook description, resulting in FWs that were too concrete (for the term signal detection theory, FWs were weak signals, life-or-death consequences, whimpering baby, detection, and experience).

FWs serves as a good review. It forces students to express the terms in their own words, actively study the material, and implicitly test their knowledge. Generating synonyms probably deepens the level of processing and semantic encoding (Craik & Tulving, 1975). Even when students are not directly engaged in play (i.e., they are waiting while the other team gives and receives clues), they can play along and test their knowledge. Playing the game with other students is fun and humorous, an important element in learning (Kaplan & Pascoe, 1977; Ziv, 1988). Additionally, the game offers opportunity for clearing up confusing topics. The instructor can listen to clues and guesses and clarify topics when necessary.

The evaluation results do not adequately describe the level of enthusiasm expressed by the students while playing the game. Although students took the activity seriously, there was much laughter and excitement while playing. However, the ill-prepared and unmotivated do not enjoy FWs. To play well, students must be familiar with the material and expend effort. It is beneficial if they review the terms in preparation for the game. The peer pressure of performing often motivates them to study. Whether the student is giving or receiving clues, FWs helps students to think differently about the material.

Results

The evaluations showed that students found the activity helpful and enjoyable. Students believed that generating the FWs helped them learn the concepts (71%; results are based on percentage of students who agreed or strongly agreed with the statement). They believed that playing FWs would serve as a good test review (75%). They wanted to use FWs to study for future exams (75%) and wanted to play it again in the future (77%). Student responses indicated that half (51%) believed that playing FWs helped them learn the concepts and few indicated that they would use FWs if it was on library reserve (26%).

Discussion

FWs is an adaptable activity. Instructors can use it to review a range of material (single chapter, units, comprehensive final, or senior exit exam) and any psychology subject. Furthermore, FWs is adaptable to class and instructor preparation time. Instructors can generate a single set of cards to use repeatedly. Alternatively, if time permits, having students generate the FWs helps them preview the material and heightens their involvement. Developing the FWs requires that students think about the ideas and generate optimal words to describe them. The task encourages students to boil down the concepts into five essential words or phrases. Student selection of FWs varied greatly; some students required additional instruction in developing quality FWs. Some students generated FWs that were well conceptualized (for the term cones, FWs were rods, color, eye, retina, and daylight). Others placed too much emphasis on the textbook description, resulting in FWs that were too concrete (for the term signal detection theory, FWs were weak signals, life-or-death consequences, whimpering baby, detection, and experience).

FWs serves as a good review. It forces students to express the terms in their own words, actively study the material, and implicitly test their knowledge. Generating synonyms probably deepens the level of processing and semantic encoding (Craik & Tulving, 1975). Even when students are not directly engaged in play (i.e., they are waiting while the other team gives and receives clues), they can play along and test their knowledge. Playing the game with other students is fun and humorous, an important element in learning (Kaplan & Pascoe, 1977; Ziv, 1988). Additionally, the game offers opportunity for clearing up confusing topics. The instructor can listen to clues and guesses and clarify topics when necessary.

The evaluation results do not adequately describe the level of enthusiasm expressed by the students while playing the game. Although students took the activity seriously, there was much laughter and excitement while playing. However, the ill-prepared and unmotivated do not enjoy FWs. To play well, students must be familiar with the material and expend effort. It is beneficial if they review the terms in preparation for the game. The peer pressure of performing often motivates them to study. Whether the student is giving or receiving clues, FWs helps students to think differently about the material.

References

In some situations, the effort expended when working alone. In some situations, the effort expended when working for social loafing often leads a group member to compensate groups was the perception of social loafing.

The major, unique feature of these negative academic work experiences and only 26% of the positive group experiences. Events associated with academic work groups comprised 54% of the negative group experiences. As expected, students in the positive past group experiences were more influential on these ratings than were current group experiences, we assigned some participants to a positive current group experience condition and others to a negative current group experience condition. We hypothesized that students whose past group experiences were negative would be less satisfied with their current group's process and performance even if the current group experience was positive, and in turn, be less willing to work on another group activity.

We hypothesized that students whose past group experiences were positive would be more satisfied with their current group’s process and performance and would be more willing to work on another group activity.


deploy collaborative learning with traditional teaching (Meyers, 1997). Between 1974 and 2002, Teaching of Psychology included 80 articles describing different collaborative strategies. Examples of such collaborative learning included group research projects, role-playing exercises, discussion groups, and debates (Meyers, 1997).

Research suggests, however, that students do not always enjoy these experiences (Phipps, Phipps, Kask, & Higgins, 2001; Wheelan & Lisk, 2000). For example, Forrest, Kershaw, and Bott (1998) asked college students to describe their most positive and negative group experiences. Events associated with academic work groups comprised 54% of the negative group experiences and only 26% of the positive group experiences. The major, unique feature of these negative academic work groups was the perception of social loafing.

Williams and Karau (1991) examined how an expectation for social loafing often leads a group member to compensate or work harder in a group than he or she would have worked alone. In some situations, the effort expended when working in a group is consistent with a “preexisting effort script” (Karau & Williams, 1993, p. 685). This script involves an individual’s expectations for how much he or she will need to contribute for the group to be successful and is based on past group experiences. In brief, students may develop negative attitudes toward group work because they have had to compensate for loafing members to obtain acceptable outcomes on group assignments. As a result, they may conclude that they will need to contribute more than their fair share in future similar projects. Accordingly, teachers need to be more cognizant of the types of group experiences that students have had, along with the effects of these experiences on satisfaction with current and future groups.

The goal of this study was to examine the extent to which students’ past as compared to current group experiences influenced ratings of group process, performance, and commitment to future group work. To investigate whether past group experiences were more influential on these ratings than were current group experiences, we assigned some participants to a positive current group experience condition and others to a negative current group experience condition. We hypothesized that students whose past group experiences were negative would be less satisfied with their current group’s process and performance even if the current group experience was positive, and in turn, be less willing to work on another group activity.

Method

Design and Measures

Our design was a 2 (past group experience: positive, negative) × 2 (current group experience: positive, negative) factorial on measures of group process, performance, and commitment to future group work.

Past group experience. To evaluate past group experience, we asked participants “In general, how do you rate your past experiences working in groups?” using a 7-point scale ranging from 1 (negative) to 7 (positive). We used a median split to categorize participants as having either positive or negative past group experiences.

Current group experience. Based on previous research (Forrest et al., 1998), our goal was to systematically create group experiences that included or did not include social loafing. In the positive current group conditions, the leader expressed interest in the task by stating that the task was interesting, maintaining eye contact with other group members who were speaking, and taking copious notes. In the negative current group condition, the leader expressed disinterest in the task by stating that the task was boring, refusing to maintain eye contact with other group members when they spoke, and taking limited notes. Regardless of the condition, group leaders were not allowed to contribute to the task solution itself.

Dependent measures. The statement “I was pleased with the way the group discussed the problem” served as the measure of group process. The statement “I was pleased with
the way the group solved the problem” served as the measure of performance. The statement “If given the opportunity, I would like to participate in another group activity” served as the measure of commitment. Participants responded to these items on a 5-point scale ranging from 1 (disagree) to 5 (agree).

Participants and Procedures

Sixty-four participants (44 women, 20 men, M = 20.7 years) enrolled at a midsize Midwestern university completed the study in return for extra credit in their introductory psychology courses. We dropped 4 participants from the sample due to incomplete data, thereby leaving 60 participants for analysis. Thirty problem-solving groups met outside of regularly scheduled class time and included 2 naive participants and 1 confederate leader. After obtaining informed consent, we randomly assigned participants to either a positive or negative current group condition. The task had group members study a circle that was divided into six wedges. Five of the wedges had letters and one was blank. The group’s task was to solve for the missing letter (see Beard, 1969). To ensure consistent leadership contributions, we directed all leaders not to help with the task but to take notes as the other group members worked on the problem. After answering questions concerning past group experiences, group members collaborated on the task before evaluating their current group experience. We debriefed participants about the purpose and our findings at the end of the semester in which we collected data.

Results

We conducted a 2 (past group experience: positive, negative) × 2 (current group experience: positive, negative) ANOVA on the measures of group process, performance, and commitment to future group work.

Past group experience significantly influenced measures of group process, $F(1, 56) = 5.46, p < .05$; group performance, $F(1, 56) = 5.08, p < .05$; and commitment to future group work, $F(1, 56) = 6.58, p < .05$. As Table 1 illustrates, participants who had positive as compared to negative past group experiences reported greater process and performance satisfaction, along with commitment to work in future groups.

There was a significant main effect for current group experience on group process, $F(1, 56) = 5.99, p < .05$. The effects of current group experience on performance, $F(1, 56) = 3.25, p > .05$, and commitment to future group work, $F(1, 56) = .10, p > .05$, were not significant. As expected, and in support of the effectiveness of current group experience, participants in the positive current group condition rated their groups’ processes as more satisfying than did participants assigned to the negative current group condition. On the other hand, no significant differences in personal satisfaction and commitment resulted based on current group condition. All means appear in Table 1. There were no significant interactions between current group conditions and past group experiences on process, $F(1, 56) = .23, p > .05$; performance, $F(1, 56) = 3.25, p > .05$; or commitment, $F(1, 56) = 2.57, p > .05$.

Discussion

The results supported the hypotheses that past group experiences significantly influence member satisfaction with their current groups’ process and performance as well as their commitment to future group work. The current group experience, however, affected only members’ satisfaction with the group process.

Because students are more likely to report academic work groups as negative experiences and past group satisfaction affects current group satisfaction, teachers need to find ways to improve the group experience in academic settings. One factor addressed previously in the literature is the amount of individual effort afforded to a group task and whether this contribution is divided equally among group members (Karau & Williams, 1993; Meyers, 1997; Williams & Karau, 1991). Instructors can improve group members’ efforts by structuring tasks so that the contributions of group members are critical to the overall group effort (Aronson, Bridgeman, & Geffner, 1978; see Aronson, 2000; www.jigsaw.org/steps.htm). Another technique that can improve the performance of individuals in groups is having students evaluate one another, thereby increasing accountability (Meyers, 1997).

Our research suggests, however, that such structuring of current group experiences may not be sufficient. Instructors need to be aware of their students’ past group experiences and students need to be aware of how those experiences may color their perceptions about group work. If group work is an integral part of a class, instructors should start the semester with a conversation in which students share their positive and negative group work experiences. This discussion helps the teacher identify which students have had negative group experiences and whether those negative experiences were related to social loafing, task structure, or accountability. During this discussion, the instructor has an opportunity to provide students with insights into effective group strategies such as (a) focusing on the group outcome beyond individual achievements, (b) being accountable as a group as well as individually, (c) working collaboratively rather than combining

### Table 1. Mean Ratings of Satisfaction With Group Process, Performance, and Commitment As a Function of Past or Current Group Experience

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process satisfaction*</td>
<td>4.60</td>
<td>3.96</td>
</tr>
<tr>
<td>Performance satisfaction*</td>
<td>4.75</td>
<td>4.18</td>
</tr>
<tr>
<td>Commitment</td>
<td>4.42</td>
<td>3.75</td>
</tr>
<tr>
<td>Process satisfaction*</td>
<td>4.62</td>
<td>3.94</td>
</tr>
<tr>
<td>Performance satisfaction</td>
<td>4.69</td>
<td>4.24</td>
</tr>
<tr>
<td>Commitment</td>
<td>4.13</td>
<td>4.04</td>
</tr>
</tbody>
</table>

Note. Ratings were based on 5-point scales ranging from 1 (disagree) to 5 (agree). *Means in rows significantly different at $p < .05$. $n = 40$. $n = 20$. $n = 30$. $Valence$
individual contributions, (d) learning and using skills related to group work, and (e) analyzing the group’s process toward an outcome (Johnson & Johnson, 1997, p. 17).

The discussion of past group experiences also allows students to hear about others’ experiences, which gives them the necessary information to reevaluate their own preexisting effort scripts as well as to determine how much effort may be appropriate for this new group. By sharing positive group work experiences, members may learn successful strategies from past groups that they can apply to current collaborations.

Because the peer learning and the teaching inherent in group work have been shown to be effective educational tools (see Hartman, 1989; Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Webb & Grib, 1967), it is important to provide students with an understanding of group processes so that they can benefit fully from such group work. Our research suggests that taking the time to neutralize the effects of negative past group experiences is essential for helping students to derive the many educational benefits that may flow from group projects.

References

their learning on this paper assignment was superior to previous writing assignments, and even fewer students reported higher motivation for the progressive writing assignment.

Indeed, the enthusiasm for progressive writing may exceed the empirical justification for it. Do the potential benefits of progressive writing result from simply receiving instructor feedback or from the multiple stages prescribed by the progressive approach? For example, a traditional two-draft (i.e., rewrite) assignment may provide instructor feedback without the fragmentation inherent in progressive writing. Furthermore, although there are clear potential benefits to the students, there are also costs to the instructor’s time in grading multiple assignments and providing feedback to students (Kalia, 1984). The purpose of this study was to investigate both the costs and benefits of progressive writing assignments.

Method

Participants

The sample consisted of 196 students taking introductory psychology courses at a Midwestern liberal arts university. The participants originated from four sections of introductory psychology taught by two instructors within 1 academic year. Each section of introductory psychology extended over the length of a semester, and the two instructors taught their sections during the same semesters, in the same classroom, 1 hr apart.

Procedure

The students completed a writing assignment similar to Hemenover et al. (1999): They viewed a popular movie and then discussed and analyzed the presentation of course topics represented in the film. There were three major sections to this assignment: movie description, psychological concept descriptions, and an analysis of the integration of the concepts within the movie. The assignment for each semester was identical with regard to content and the criteria for grading; the only difference was whether the students turned in the paper by sections (i.e., progressive writing) or in its entirety as the first draft (i.e., traditional two-draft format). Students used identification numbers, not names, on their papers to avoid any potential for instructor expectancy bias. With both writing formats, students responded to the instructors’ written feedback, had the opportunity to meet with the instructor, and made changes before the final paper grade was assigned. Each instructor acted as her own control by offering the progressive writing assignment in one semester and the rewrite assignment in the other semester. The order of presentation (two-draft vs. progressive) was counterbalanced to avoid any potential for instructor bias.

Materials

This study incorporated several measures of the students’ perceptions of the writing assignment. At the beginning of the semester, after the explanation of the paper assignments, students indicated on a 10-point scale how clear the directions for writing paper were (1 [very unclear, I am very confused] to 10 [very clear, I know exactly what to do]) and how anxious they were about the paper assignment (1 [not at all anxious] to 10 [very anxious]). At the midpoint in the semester, students again indicated how clear the directions were and how anxious they were about the writing assignment.

At the end of the semester, after the completion of the paper assignment, we again asked students how clear the directions were and how anxious they were about the assignment. Additionally, students indicated on 10-point scales how much they had learned about psychology through this assignment (1 [nothing] to 10 [learned a lot]) and how much their writing had improved with this assignment (1 [not at all] to 10 [very much]). Students then read a brief description of a traditional two-draft rewrite assignment and of a progressive writing assignment, and they indicated which type of assignment they would have preferred if they had a choice. Moreover, this study included the students’ final paper grades as an instructor-derived measure of the quality of writing. Finally, we tracked the total number of hours we spent on paper-related activities (e.g., grading, meeting with students) over each semester as an indicator of the time required by each format.

Results

To measure any potential differences among the two groups’ abilities, analyses compared the progressive writing group’s average test performance with the two-draft writing group’s test scores. Students in the progressive writing condition scored 3 percentage points higher than their traditional two-draft peers, F(1, 194) = 8.40, p < .01. Although not practically significant for this study because 3 percentage points did not result in a grade difference using our grading scale, all of the remaining analyses controlled for any differences in exam scores.

Student Perceptions

Repeated measures ANCOVAs were computed with time (prettest, midsemester, and posttest) as a within-subjects variable, format (progressive, two-draft) as a between-subject variable, and test performance as the covariate to assess student ratings of clarity regarding the directions as well as anxiety about completing the assignment. Results demonstrated a significant Time × Format interaction concerning the clarity of directions for the writing assignment, F(2, 192) = 4.05, p < .05. Follow-up tests revealed that students with the progressive format were more confused midway through the semester than their counterparts, t(194) = –3.72, p < .001 (see Table 1). There were no significant differences in anxiety over the course of the semester for either format.

Similar to Hemenover et al. (1999), univariate ANCOVAs with test performance as a covariate revealed no significant differences between the paper formats in the students’ end-of-the-semester perceptions about their learning of psychology or their improvement in writing ability as a result of the paper assignment. Finally, students preferred the assignment version they experienced, χ²(2, N = 104) =
Table 1. Measures of Central Tendency and Variability for Progressive and Two-Draft Writing Formats

<table>
<thead>
<tr>
<th></th>
<th>Progressive Format</th>
<th>Two-Draft Format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Beginning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of assignment directions</td>
<td>7.38</td>
<td>1.45</td>
</tr>
<tr>
<td>Anxiety about writing assignment</td>
<td>5.60</td>
<td>2.12</td>
</tr>
<tr>
<td><strong>Midsemester</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of assignment directions</td>
<td>7.19***</td>
<td>1.82</td>
</tr>
<tr>
<td>Anxiety about writing assignment</td>
<td>5.90</td>
<td>2.03</td>
</tr>
<tr>
<td><strong>End of semester</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of assignment directions</td>
<td>7.30</td>
<td>2.05</td>
</tr>
<tr>
<td>Anxiety about writing assignment</td>
<td>5.82</td>
<td>2.25</td>
</tr>
<tr>
<td>Percentage of students who preferred their assigned writing format</td>
<td>64.42</td>
<td>67.42</td>
</tr>
<tr>
<td>Amount learned about psychology</td>
<td>6.87</td>
<td>1.89</td>
</tr>
<tr>
<td>Improvement in writing skills</td>
<td>5.64</td>
<td>2.20</td>
</tr>
<tr>
<td>Paper grade (out of 100)</td>
<td>88.71**</td>
<td>8.31</td>
</tr>
<tr>
<td>Amount of time to grade assignment (in hr)</td>
<td>66.59*</td>
<td>7.90</td>
</tr>
</tbody>
</table>

*ns = 105, n = 91. Anchors: 1 (very unclear, I am very confused) to 10 (very clear, I know exactly what to do). Anchors: 1 (not at all anxious) to 10 (very anxious). Anchors: 1 (nothing) to 10 (learned a lot). Anchors: 1 (not at all) to 10 (very much).

*p < .05. **p < .01. ***p < .001.

60.94, p < .001 for the progressive format and $\chi^2(2, N = 89) = 57.06, p < .001$ for the two-draft format.

Instructor-Derived Data

A univariate ANCOVA with test performance as a covariate revealed that students assigned to the two-draft format scored slightly higher on the paper, $F(1, 195) = 9.43$, $p < .01$. The average grade difference between the progressive and draft formats was less than 3 percentage points (see Table 1). As expected, the progressive version of the writing assignment took significantly longer to grade, $t(2) = 3.79$, $p < .05$. In fact, the progressive writing version of the assignment took 45.5% more total time to grade than the two-draft writing format.

Discussion

The results of this study indicated that the progressive format did not yield a better product as evidenced by grades, student reports of learning, student perceptions of improvement in writing ability over the course of the semester, clarity, student anxiety, or student preference. The one clear difference in this study was that progressive writing required significantly more grading time for the instructor.

It is important to emphasize that empirical research on progressive writing is essentially nonexistent. A need exists for future researchers to further investigate the merit of progressive writing. The type of paper (e.g., application of concepts vs. literature review vs. generative research), the length of the assignment, the discipline (e.g., psychology vs. journalism), and many other factors may impact the utility of progressive writing as a pedagogical technique. Yet, the overall message of this study is that instructors need not necessarily abandon assigning traditional two-draft papers given that there were not demonstrated advantages of the grading-intensive progressive writing format.

References


Note

Send correspondence to Laura L. Finken, Department of Psychology, 2500 California Plaza, Creighton University, Omaha, NE 68178; e-mail: lfinken@creighton.edu.

General Psychology Course Evaluations: Differential Survey Response by Expected Grade

Elizabeth M. Ginexi
George Washington University

This study used general psychology course evaluation survey data to test whether all ratings are uniformly associated with expected grade or whether an expected grade–rating relation exists only for certain questions. Students rated the course difficulty and how much they liked the textbook and the teaching. Students who expected higher grades provided more favorable teacher ratings, and they found the assigned readings easier to comprehend. Expected grade was completely unrelated to other responses, including the number of study hours and liking the textbook. More research is needed on the kinds of questions affected by expected grade effects and on possible explanations for differential effects. Research on
impression management theory and exit interviewing may offer one promising approach.

For decades student evaluations of teaching have been a popular measure of performance for college and university instructors, but their use has not been without criticism. One literature search indicated that there are thousands of papers on the topic (Marsh & Dunkin, 1992). Researchers have examined numerous influencing factors such as timing of administration (Seldin, 1989), location of administration (Munz & Fallert, 1998), anonymity of responses (Blunt, 1991), student or instructor gender (Bennett, 1982; Tatro, 1995), instructor personality (Radmacher & Martin, 2001) and, perhaps most controversially, students’ expected grade. There appears to be a moderate positive correlation between expected grade and student ratings (for a review, see Wachtel, 1998).

Marsh and Roche (1997) summarized two popularly competing interpretations. The grading-leniency hypothesis posits that students give higher ratings to instructors who assign higher than deserved grades (e.g., Greenwald & Gillmore, 1997). The validity hypothesis proposes that better expected grades reflect better learning by students (e.g., Peterson & Cooper, 1980). Although the validity hypothesis does not explicitly specify how teacher ratings may be affected, it seems likely that perceived mastery level could affect ratings of course difficulty.

Investigations into these competing interpretations of the expected grade–teacher rating relation have neglected to consider the unique context under which students provide evaluation responses and the role this context may play in explaining the association. Social psychologists and survey methodologists have long been aware that attitude assessment is highly context dependent (Eagly & Chaiken, 1993; Schwarz, 1999). When surveys are voluntary, the motivation to respond may be one of the most important contextual factors.

Research on the motivational and cognitive aspects of exit interviews and surveys suggests that people often distort responses to exit surveys to leave a particular image of themselves (Giacalone & Duhon, 1991). In organizational contexts, employees who voluntarily leave a job may be motivated to maintain a favorable impression of themselves, and therefore, they are likely to provide positive or neutral responses in exit interviews. On the other hand, employees who leave a job involuntarily are experiencing a negative event, and as a result, their desire to manage a favorable impression may be replaced by a desire to create an impression of the organization that is negative or troublesome (e.g., Giacalone, Knouse, & Ashworth, 1991; Giacalone & Rosenfeld, 1991). Organizational survey researchers have found that such retaliation-based responses are frequently focused on more unverifiable issues such as supervisor quality, rather than on more readily verifiable issues such as available medical benefits or office equipment (Giacalone, Elig, Ginexi, & Bright, 1995; Giacalone & Rosenfeld, 1991).

Students exiting a course with the expectation of a low grade may respond to an evaluation survey from a negative perspective similar to one espoused by a disgruntled worker. Based on prior exit survey research, it seems likely that the exiting circumstances, in this case expected grade, may affect certain question content areas more than others. This article provides a preliminary investigation of whether all student ratings are uniformly associated with expected grade or whether expected grade–rating relations are observed only for certain questions.

Testing for differential response patterns as a function of expected grade might offer some clues about the mechanisms driving the expected grade–rating association. If grading leniency is the main motivator for the relation between expected grades and survey responses, the bias might prompt students with higher grades to provide uniformly favorable ratings. To the extent that the validity hypothesis rings true, one might find that expected grades are correlated primarily with students’ beliefs about mastery or course difficulty. Finally, an impression management interpretation may be supported if the expected grade–rating relation mostly affects questions pertaining to impressions of the instructor.

Method

Participants and Procedure

Students in a general psychology course participated in the evaluations presented here. The course, which took place in the fall of 1997, had 136 students enrolled. The evaluation survey was both voluntary and anonymous. Students who were willing to participate picked up a blank survey form after turning in their final exam. As per the instructions, students did not put their names or any other identifying information on the forms. A total of 105 students completed the survey (response rate = 77%). The survey questions appear in Table 1.

Results

Expected grade was represented by a three-level variable indicating grades of A, B, or C/D. Effects of expected grade on each question in Table 1 were tested with seven ANOVAs. To adjust for potential inflation of the Type I error rate, the results were interpreted with a more conservative alpha criterion of .007. Table 2 lists the ANOVA results, means by expected grade, and a posteriori contrasts.

Only three models reached significance at the conservative alpha criterion. Students with higher expected grades found the textbook easier to comprehend, and students who expected to receive higher grades provided more favorable teacher ratings. Expected grade was unrelated to the number of hours studied per week and liking the assigned textbook.

Discussion

The grading-leniency hypothesis seems unlikely to have played a role in these findings because only certain questions were affected by an expected grade–rating correlation. The

Vol. 30 No. 3, 2003 249
validity hypothesis gained some support in that students who believed they best understood the text also expected higher grades. Notably, the expected grade–rating relation was observed mostly with respect to questions about the instructor and not for questions about study effort and liking the assigned textbook. This latter finding points to the possibility that student evaluations of teaching may reflect retaliation-based responses similar to those observed in exit surveys. Students who expect to receive a high grade may be motivated to leave favorable impressions, but students who expect to receive lower grades may be motivated to get back at the instructor by offering lower teacher ratings. However, expected grades do not seem to motivate extreme ratings on topics that are less relevant to instructor reputation such as study hours or textbook popularity. Indeed, past research on college students’ perceptions of the evaluation process indicates that they may view evaluations as a way to vent or let off steam (Marlin, 1987).

The study’s limitations must be mentioned. There are obvious problems with generalizability to other courses and instructors. The sample size was small (N = 105), and it is unclear whether the students in this class were representative of students who regularly enroll in general psychology courses. In addition, not all students chose to complete the survey. It is unclear whether the findings might have differed for the students who did not participate. Also, it is possible that other, unmeasured variables could account for the relations found. For example, some studies suggest that evaluations reflect students’ assessments of instructor personality (e.g., Radmacher & Martin, 2001). Clearly, these results are only preliminary findings. Nevertheless, this study demonstrates that expected grade effects may not be uniformly ob-

Table 1. Survey Questions by Content Area and Descriptive Statistics

<table>
<thead>
<tr>
<th>Question</th>
<th>n</th>
<th>M/Freq</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. How easy/difficult to comprehend was the Gleitman text? (0 [very difficult] to 7 [very easy])</td>
<td>105</td>
<td>4.46</td>
<td>1.42</td>
<td>1 to 7</td>
</tr>
<tr>
<td>2. How many hours a week would you estimate that you studied/read for this course?</td>
<td>94</td>
<td>3.95</td>
<td>2.63</td>
<td>0 to 15</td>
</tr>
<tr>
<td>Like the textbook</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. How interesting was the Gleitman text, Basic Psychology? (0 [not at all interesting] to 7 [very interesting])</td>
<td>104</td>
<td>4.36</td>
<td>1.47</td>
<td>0 to 7</td>
</tr>
<tr>
<td>4. Would you recommend that I use this text again, or should I find another one? (0 [definitely find another text] to 7 [definitely use again])</td>
<td>105</td>
<td>4.68</td>
<td>1.97</td>
<td>0 to 7</td>
</tr>
<tr>
<td>Evaluation of teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Did you find the class demonstrations interesting? (0 [very uninteresting] to 7 [very interesting])</td>
<td>104</td>
<td>4.48</td>
<td>1.79</td>
<td>0 to 7</td>
</tr>
<tr>
<td>6. Would you consider taking other courses if I taught them? (0 [definitely not] to 7 [definitely yes])</td>
<td>101</td>
<td>4.08</td>
<td>2.09</td>
<td>0 to 7</td>
</tr>
<tr>
<td>7. Would you recommend me as an instructor to other students? (0 [definitely not] to 7 [definitely yes])</td>
<td>100</td>
<td>4.42</td>
<td>1.97</td>
<td>0 to 7</td>
</tr>
<tr>
<td>Expected grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. At this point, what grade do you expect to receive in this course (A, B, C, D)?</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>21</td>
<td>21.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>45</td>
<td>46.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>26</td>
<td>27.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>4.2%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. A plus sign (+) denotes means that are listed for descriptive purposes only because the univariate tests were not significant at the criterion .007 alpha level. In each row, means with different letter subscripts are significantly different statistically, based on Duncan’s multiple-range tests. Level of statistical significance for all contrasts is p < .05.

Table 2. ANOVA Tests, Means by Expected Grade, and a Posteriori Contrasts

<table>
<thead>
<tr>
<th>Question</th>
<th>F</th>
<th>df</th>
<th>p &gt; F</th>
<th>Expected Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. How easy/difficult to comprehend was the Gleitman text?</td>
<td>5.37</td>
<td>2</td>
<td>.006</td>
<td>5.10&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>2. How many hours a week would you estimate that you studied/read for this course?</td>
<td>0.25</td>
<td>2</td>
<td>.787</td>
<td>4.14+&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>Like the textbook</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. How interesting was the Gleitman text, Basic Psychology?</td>
<td>0.94</td>
<td>2</td>
<td>.396</td>
<td>4.38+&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>4. Would you recommend that I use this text again, or should I find another one?</td>
<td>1.97</td>
<td>2</td>
<td>.145</td>
<td>4.90+&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>Evaluation of teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Did you find the class demonstrations interesting?</td>
<td>4.18</td>
<td>2</td>
<td>.018</td>
<td>5.14+&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>6. Would you consider taking other courses if I taught them?</td>
<td>7.21</td>
<td>2</td>
<td>.001</td>
<td>5.24&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>7. Would you recommend me as an instructor to other students?</td>
<td>7.46</td>
<td>2</td>
<td>.001</td>
<td>5.52&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Note. A plus sign (+) denotes means that are listed for descriptive purposes only because the univariate tests were not significant at the criterion .007 alpha level. In each row, means with different letter subscripts are significantly different statistically, based on Duncan’s multiple-range tests. Level of statistical significance for all contrasts is p < .05.
served across different evaluation questions. Future research on associations between expected grade and evaluations of teaching should examine in more depth what question content areas are most likely to be affected and explore possible explanations for differential question effects. One promising approach is to examine student evaluations within the context of the existing knowledge of impression management theory and exit interviewing.

References


Note

Send correspondence to Elizabeth M. Ginexi, who is now at Westat, 1650 Research Boulevard, Rockville, MD 20850; e-mail: ElizabethGinexi@Westat.com.

A Psychic-Reading Demonstration Designed to Encourage Critical Thinking

Timothy J. Lawson
College of Mount St. Joseph

Self-proclaimed psychics who perform seemingly accurate readings have been the focus of a number of recent television shows. I describe a psychic-reading demonstration—in which a professor performs a surprisingly accurate reading with a student in class—designed to encourage students to think critically about such readings. Students reported that they found the demonstration to be interesting, informative, and helpful for thinking about alternative explanations for the apparent accuracy of psychic readings.

People’s strong interest and long-standing belief in questionable pseudoscientific and paranormal phenomena have prompted a number of psychologists to develop strategies for using such phenomena to teach students critical thinking skills (e.g., Banziger, 1983; Bates, 1991; Lilienfeld, Lohr, & Morier, 2001; McBurney, 1976; Wesp & Montgomery, 1998). These authors described ideas ranging from classroom demonstrations of telepathy to entire courses on pseudoscientific and paranormal topics.

Recently, a number of self-proclaimed psychic readers (e.g., Sylvia Browne, John Edward, James Van Praagh), some of whom supposedly communicate with the dead, have been the subject of a number of television shows and magazine articles (e.g., Christopher, 2001; Jaroff, 2001; Wiseman & O’Keefe, 2001). Seeing a reader reveal accurate and sometimes detailed personal information about a target person who is apparently a stranger may lead observers to infer that the reader is psychic. To encourage students to think critically about such readings, I developed a psychic-reading demonstration, including follow-up discussion questions. Although previous authors (e.g., Bates, 1991; McBurney, 1976) have described ideas for demonstrating telepathic abilities in the classroom (e.g., by guessing the playing cards or numbers selected by students), they did not describe how to perform a psychic reading that involves revealing specific details about a person one does not know well.

My demonstration relies on “cold-reading” (e.g., making general statements that apply to most people) and “hot-reading” (e.g., obtaining specific details about the person in advance) techniques similar to those apparently used by famous readers (see Jaroff, 2001; Nickell, 2001; Randi,
Because of the ethical problems associated with a reading focused on deceased friends or relatives (see Nickell, 2001), I focus the reading on the person’s personality and past experiences (as do many psychic readers). Consistent with the recommendation of Seckel (1989), the main purpose of the demonstration is not to simply debunk the readings performed by famous readers by explaining how they might be performed; it is to encourage critical thinking by asking students to generate alternative explanations and determine whether accurate, detailed readings constitute high-quality evidence for the reader’s psychic powers.

The Psychic-Reading Demonstration

I perform the demonstration in my introductory psychology class about 5 weeks into the semester, during our discussion of extrasensory perception. I state that I am going to do a psychic reading and need a “volunteer.” I act as if I select a student arbitrarily (e.g., “Let’s get someone who sits in the front row”), but I make my choice prior to this class. I ask the student to come to the front of the room and hand me a personal possession, such as car keys. I put the keys in my hands, and concentrate intently. Then, I begin telling the student what I “see” about his or her personality. For this part of the reading, I make several general statements that should apply to most people, such as “You are outgoing at times, but reserved at other times,” “You are fairly even-tempered, but sometimes get very angry,” and “You enjoy helping others.”

Afterward, I slowly disclose detailed information about the student, acting as if bits and pieces of information are “coming” to me. I pause and act uncertain at a number of points during the reading, and often start with vague information before being more specific (e.g., “I see the letters WR, what does that mean? Did you play wide receiver?”). During one reading I stated that the target student liked basketball, played guard on his high school basketball team, played a game in Florida against a team named Miami Christian, and scored five 3-point shots and 23 total points during that game. This student was so surprised that he jumped out of his chair twice during the reading. In another reading, I disclosed that one of my students grew up in a single-parent household, was the captain of his high school’s cross-country team, won a Burger King A+ Award for his cross-country achievements, and suffered a broken leg when he was hit by a truck at a young age. The class was shocked when I revealed these details during my reading.

Prior to the readings, I obtain detailed information about one or two of the students in the class. Good, public sources of information include student introductions during the first day of class, marriage announcements, student Web pages, and local newspaper articles on students. The majority of the information I use comes from articles I find on the World Wide Web. I have found it is easiest to find information about student athletes because their names seem most likely to appear in the media. To avoid an unethical invasion of students’ privacy, I am careful to use only information that the target students know is publicly available.

After the reading, I distribute a handout to the students and ask them to form small groups to answer the critical thinking questions on the handout. The questions ask them to determine (a) if a target person’s acknowledgment of the accuracy of a reading constitutes high-quality evidence for the psychic abilities of the reader, (b) whether there are alternative explanations for the accuracy of the reader’s statements, and (c) how they might design a test to determine if a reader is able to obtain information about a person by psychic means.

I tell the class that I am not psychic, and we discuss their answers to the questions. Typically, some students state that readers might make vague statements that should apply to most people, and I briefly explain cold-reading techniques. Some students also state that readers might obtain information about the target person in advance of the reading, and we discuss examples of hot-reading techniques that might be used by famous readers. Afterward, I explain how I obtained the detailed information for my reading. While discussing how to test a psychic, one suggestion I make is to ask the psychic several specific questions on topics about which the psychic would have no information (e.g., the student’s favorite high school teacher). Finally, I also recommend informing students that the search for student information (a) was restricted to one or two students, (b) included only publicly available sources, and (c) was performed solely for the purpose of the demonstration. Because this demonstration requires temporarily deceiving students by pretending to be psychic, professors might also discuss ethical issues surrounding the use of deception.

Students’ Reactions to the Demonstration

Seventy-three students enrolled in one of three sections of introductory psychology indicated—on a scale ranging from 1 (strongly disagree) to 5 (strongly agree)—that they found the demonstration to be interesting (M = 4.42, SD = 0.62) and informative (M = 4.18, SD = 0.63). They also agreed that the demonstration would help them think about alternative explanations for “accurate” psychic readings in the future (M = 4.25, SD = 0.94). Students’ written comments about the demonstration were also very positive. They described it as “convincing,” “cool,” “fun,” “very informative,” and “really interesting.” A number of students said they were shocked by my accuracy, and one said, “We really thought you had psychic abilities at first, until you explained.” However, 2 students indicated that they were suspicious because I had “too much information.” In short, I have found this exercise to be an engaging, entertaining way to encourage students to think critically about psychic readings. In fact, of the many demonstrations I use in my classes, I believe this one generates the most student interest and excitement.

References


I examined research methods, child development, and developmental psychology textbooks (N = 74) for ethical considerations concerning child research participants. Coverage varied both within and between textbook types, with many texts showing little or no discussion, especially of debriefing and recruiting incentives. Greater textbook attention to child participants would benefit students by raising their awareness of ethical considerations designed to safeguard children’s rights and welfare.

Educating psychology majors about research ethics is an essential component of the undergraduate curriculum (Brewer et al., 1993). Accordingly, course instructors have developed teaching methodologies that introduce students to ethical principles and teaching techniques to integrate this information into psychology course content (e.g., Beins, 1993; Matthews, 1991). Teachers of psychology also have investigated textbook coverage of research ethics because textbooks are a significant source of information for students (McKeachie & Hofer, 2001). Two early studies revealed that coverage of research ethics was missing in 44% (n = 18) of experimental (Adair, Lindsay, & Carllopio, 1983) and 21% (n = 29) of introductory psychology textbooks (Korn, 1984). More recently, however, Fisher and Kuther (1997) reported that each of 14 introductory psychology textbooks mentioned research ethics. Nonetheless, results from introductory textbooks may not generalize to other psychology textbooks, and none of the studies previously mentioned explicitly addressed research ethics concerning child participants.

Federal laws and ethical codes promulgated by professional organizations mandate that investigators protect all research participants from physical and psychological harm. Researchers, therefore, should apply these standards of conduct to evaluate the ethical acceptability of their research. Ethical codes, however, are not exhaustive standards of conduct; instead some ethical codes complement each other. Thompson (1990) and Miller (1998) asserted that the Ethical Standards for Research With Children (Society for Research in Child Development [SRCD], 1993) provide a more sensitive appraisal of children’s vulnerabilities than the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association [APA], 1992). Thus, it is common practice to apply more than one set of ethical guidelines when planning and conducting research (Miller, 1998).

Haemmerle and Matthews (1988) suggested that coverage of ethical principles should occur in at least one required psychology course, although undergraduate programs have not standardized which course satisfies this recommendation. The research methods (RM) course seems best suited to fulfill this recommendation because it is a required course of most psychology majors (Messer, Griggs, & Jackson, 1999), and it is more directly concerned with research ethics than other courses in the undergraduate curriculum. Moreover, social scientists consider research ethics a necessary topic of the RM course (Giesbrecht, Sell, Scialfa, Sandals, & Ehlers, 1997). As such, expecting RM textbooks to include ethical considerations for child participants seems plausible. One also might expect to find coverage of child participants in undergraduate textbooks for child development (CD) and developmental psychology (DV) courses. To date, however, textbook coverage of child participants has received no attention. Thus the purpose of this investigation was to determine the extent to which RM, CD, and DV textbooks present ethical considerations in research with children.

**Method**

*I obtained textbooks copyrighted from 1995 through 1999 from colleagues and publishers. The time frame matched other current textbook research (e.g., Jackson, Lugo, & Griggs, 2001), and it allowed the inclusion of the most recent textbooks available at the time of this study. Excluded were textbooks published outside the United States, combined RM and statistics textbooks, student laboratory manuals, handbooks, study guides, and instructor’s manuals. The sample (N = 74) resulted in 30 RM, 27 CD, and 17 DV textbooks (references for the 74 texts are available on request).*
estimated the sample represents 95% of the textbooks published within each text category.

Coding System

The coding system focused exclusively on ethical considerations of children's participation in nonclinical research. Not included were ethical concerns related to therapeutic interventions. The coding system consisted of two parts. One part coded textbook information for the presence of the APA Ethics Code (APA, 1992) and the SRCD Ethical Standards (SRCD, 1993). This part also accounted for any citations or references to other professional ethical guidelines and standards and to journal articles, books, and book chapters that address child participants in any way.

The second part of the coding system consisted of six topics that warrant particular consideration with child participants. These provisions were the most frequently cited ethical considerations with child participants derived from a preliminary examination of a random sample of 50% of the textbooks in each text category. The special provisions that emerged were (a) parental or guardian permission, (b) child assent, (c) freedom to decline or withdraw from participation, (d) confidentiality, (e) recruitment incentives, and (f) debriefing children. The coding system counted each provision as present if the textbook explicitly included children or adolescents in a statement, narrative description, or discussion of it. Additionally, it counted an example of a research study, either hypothetical or actual, that illustrated one or more of the six provisions. This way, a research study could substitute for a provision not explicitly mentioned. Finally, the coding system included the occurrence of a developmental portrayal of research risk.

Procedure

Two research assistants searched the table of contents and the subject index of each textbook for sections addressing child participants, ethical principles, and the special provisions. For selection of ethics-relevant pages, interrater agreement exceeded .97. Each assistant then read in detail and, in random order, coded half the ethics-relevant pages in each text category, with 50% overlap between them to assess reliability. Coding agreement across all topics exceeded .93. The assistants resolved coding disagreements through discussion.

Results

Table 1 displays the number and percentage of RM, CD, and DV textbooks that included the APA Ethics Code, the SRCD Ethical Standards, and citations of other professional codes of conduct and ethics resources. Coverage of the APA and SRCD ethical guidelines in CD and DV textbooks showed a similar pattern, although three CD (11%) and four DV (24%) texts included neither set. More striking is that all 30 RM textbooks included the APA Ethics Code, but only 4 RM texts (13%) covered the SRCD Ethical Standards.

Table 1. Number and Percentage of Textbooks Addressing Selected Topics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Research Methods</th>
<th>Child Development</th>
<th>Developmental Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>APA Ethics Code</td>
<td>30</td>
<td>100</td>
<td>19</td>
</tr>
<tr>
<td>SRCD Ethical Standards</td>
<td>4</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>Both ethical guidelines</td>
<td>4</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Neither ethical guidelines</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other professional codes</td>
<td>6</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Research ethics resources</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. The percentage of textbooks in each category was rounded to the nearest whole percent.

Across textbook types, the overall percentage of texts that included other professional codes of conduct or ethical guidelines was quite low (20%). Only one RM and one CD textbook cited references to published works concerning ethical research with children (see Table 1). Combined, these two textbooks provided the following references: Fisher, Higgin-D’Alessandro, Rau, Kuther, and Belanger (1996); Koocher and Keith-Spiegel (1990); and Thompson (1990).

Table 2 displays the special provisions and respective percentages for each textbook type. Of the six provisions, CD texts covered four, DV texts covered three, and RM texts covered only two at more than a 40% frequency. Yet the occurrence of all six provisions in a single textbook was unusual—only one CD textbook included all of them. The inclusion of a research study as an illustrative example of the special provisions occurred in only 10 CD (37%), 9 RM (30%), and 4 DV textbooks (24%).

The overwhelming majority of textbooks failed to include a developmental portrayal of research risk. Specifically, 23 CD (85%), 25 RM (83%), and 14 DV (82%) textbooks were silent on this topic. Those textbooks that did address children’s vulnerabilities to research risk typically described investigative situations to highlight developmental differences.

Discussion

Two noteworthy findings emerged from this investigation. First, across textbook types, coverage of the APA Ethics Code and the SRCD Ethical Standards is disproportionate, with the greater disparity occurring in RM textbooks. One explanation for this finding is that RM textbook authors are more familiar with the APA Ethics Code than the SRCD Ethical Standards. Another explanation is that, compared
with the SRCD Ethical Standards, the APA Ethics Code is broader in scope, addressing many areas of professional ethical conduct. Thus, textbook authors who have space limitations might be forced to choose between the two sets of ethical guidelines. Also, to appeal to a broad audience, RM textbook authors may have limited or omitted coverage of child participants, but included ethical concerns related to therapeutic interventions. Because this study excluded coverage of children’s participation in clinical trials and biomedical research, the data do not reflect these considerations.

The second notable finding is the paucity in textbook coverage of the special provisions. Specifically, many textbooks either neglected parental permission and child assent or included parental permission and ignored child assent. Consequently, these omissions might easily lead students to two incorrect conclusions: First, parental permission would supercede child assent, when in fact, children have the right to decide whether to participate (see 45 CFR 46.402 Subpart D.; Department of Health and Human Services, 1991). Second, children’s failure to dissent indicates agreement to participate. Because parents and children both have a voice in the decision to participate (Tymchuk, 1992), each warrants attention. Textbooks ignoring the topic of child assent suggests an insensitivity to children’s rights.

With few exceptions, RM textbooks are not an adequate source for the ethical treatment of child participants. Unfortunately, some CD and DV textbooks fail to fill this void. In addition, undergraduate psychology programs are offering child development courses less frequently today, with greater emphasis on developmental psychology and life span courses (Perlman & McCann, 1999a, 1999b). Consequently, students may have fewer opportunities to take a child development course. If the most comprehensive coverage of child participants only occurs in a few CD textbooks, it seems unlikely that students will gain exposure to the ethical considerations in research with children unless teachers fill the void.

The results of this study underscore the need for RM, DV, and CD textbooks to improve coverage of child participants. Indeed, students might overlook ethical concerns as problems because they are unaware a problem exists (Johnson, 1991; Kallgren & Tauber, 1996). To increase students’ exposure to the ethical treatment of children, I urge textbook authors to provide comprehensive coverage of child research participants. Few, if any, published teaching activities that concern research ethics explicitly include child participants. Thus, course instructors should seek to develop teaching activities to increase students’ awareness of and sensitivities to child participants. The implications of this study for textbook authors and teachers of psychology are important not only because they affect students’ knowledge of research ethics, but because they ultimately affect the ethical treatment of children.

References


Motion Parallax: Is It Presented Accurately in Textbooks?

Yancy B. McDougal
George W. Crowe
Sean M. Holland
University of South Carolina Spartanburg

We examined the presentation of motion parallax in introductory psychology and sensation and perception textbooks. We found that textbook discussions of motion parallax are almost always incom-
tailed drawings of the retinal flow during motion parallax; the drawings were important because they revealed that illustrations in other textbooks often were inaccurate. Now this investigation had taken shape: How complete and accurate are explanations of motion parallax in introductory psychology and sensation and perception texts?

In a content analysis similar to that by Kilmartin and Dervin (1997), we surveyed the presentation of motion parallax in all of the introductory psychology textbooks (1987 to 2002; n = 31) and sensation and perception textbooks (1997 to 2002; n = 6) belonging to psychology faculty members at the University of South Carolina Spartanburg. We excluded duplicates (older editions) of the introductory textbooks (n = 5) and the sensation and perception textbooks (n = 4) from the study. We also surveyed the visual perception texts (1950 to 1998; n = 10) available at the University of South Carolina Spartanburg library. Using the key words visual perception, we found more than 100 text titles in the university’s online card catalog. From that list, we selected the texts that seemed likely to address motion parallax. To assess the viability of journal articles as a source for an explanation of motion parallax, we found 177 journal article titles through PsycINFO using the key words motion parallax and limiting the search to the years between 1950 and 2001. From that list, we surveyed all of the articles that were available in the University of South Carolina Spartanburg library (1983 to 2001; n = 28). We included in the survey an article on the history of monocular depth perception, bringing the article count to 29.

Only two of the visual perception texts (i.e., Gibson, 1950; Haber & Hershenson, 1980) provided a complete (and correct) explanation of motion parallax, which they referred to as motion perspective. (Haber & Hershenson, 1980, provided a clear distinction between the terms motion parallax and motion perspective.) However, because the focus of those authors is on what Haber and Hershenson (1980) considered to be the more general cue, motion perspective, the explanations are lengthy, and the reader must combine material from several pages to derive the complete explanation. The remaining visual perception texts provided either a very brief explanation or none at all. As we expected, the journal articles provided either a very brief explanation of motion parallax (n = 12) or none at all. We also found problems with the explanations and the illustrations of motion parallax presented in the textbooks. In all texts that addressed the phenomenon, the explanation of motion parallax was incomplete, or unclear at the very least, and more than half of the illustrations of motion parallax were wrong.

Of the 31 introductory texts, 5 did not address motion parallax, and the remaining 26 provided an explanation that was incomplete, unclear, or both. The most common explanation (n = 21) was that objects closer to the observer seem to move past more quickly or seem to move a greater distance than do objects farther from the observer. Although this statement is not wrong, it is incomplete because it addresses only a horizon fixation situation. Several texts made the vague statement that objects at different distances appear to move at different speeds. Five texts provided an inconsistent explanation by addressing an intermediate fixation point along with a single, decreasing speed gradient for more distant objects. Several texts failed to mention a fixation point and a direction of apparent movement, whereas others only implied the direction of apparent movement using phrases such as “whip past” and “fly by.” Although none of the texts provided an incorrect explanation of motion parallax, none provided a clear and complete explanation.

In addition to the (typically) brief explanation of motion parallax, textbook authors sometimes provided an illustration depicting the apparent relative motion of objects as a person looks out the side window of a moving vehicle. Drawings depicting motion parallax usually contain vectors that represent the magnitude and direction of apparent movement (Ostdiek & Bord, 1995). Of the 26 introductory textbooks that addressed motion parallax, 17 had no illustrations of motion parallax, and the remaining 9 provided illustrations that were either inaccurate, incomplete, or unclear. The most common inaccuracy was to include vectors of equal length regardless of the distance of the object from the observer or from the fixation point. In some drawings, the vectors varied in length but the length was inconsistent with distance information. Another inaccuracy was to indicate direction of apparent movement reversing as if there were an intermediate fixation point but then failing to include the fixation point. In some texts, photographs (taken from the window of a moving train) contained no vectors and no fixation point. One text provided an accurate drawing from Haber and Hershenson’s (1973) text; however, the authors neglected to adequately explain the drawing. Finally, one text included an optical flow or expansion pattern drawing similar to Gibson’s (1950) motion perspective when looking ahead. The use of the expansion pattern is not incorrect, but it is difficult to understand relative motion in such an illustration.

Of the six sensation and perception texts surveyed, none contained an explanation of motion parallax that was complete, correct, and clear. For example, all six texts described the apparent speed of objects nearer the observer as faster than objects farther away. However, four of these texts also mentioned the direction of apparent movement reversing around an intermediate fixation point, which is inconsistent with the speed gradient. One text addressed an intermediate fixation point but not its associated speed gradient. Another text failed to address an intermediate fixation point or the change in direction of apparent movement around that point. As with the introductory psychology texts, none of the explanations included incorrect statements, but none was a comprehensive explanation of motion parallax.

Only four of the sensation and perception texts provided a drawing depicting motion parallax, and two of those drawings were inaccurate. Both drawings indicated the direction of apparent movement reversing around an intermediate fixation point. However, one drawing included vectors that were the same length regardless of the distance from fixation; in the other drawing, the vectors indicated a decreasing speed gradient from the observer to the horizon.

Our survey results indicate that textbook explanations of motion parallax are often incomplete or unclear (or both) and that complete explanations can be lengthy. In this article, we offer a brief yet clear and complete explanation of motion parallax. For a relatively simple demonstration of motion parallax, place three objects labeled A, B, and C (A being closest to you) in a line on a table or down a hallway. Stand in front of the line of objects, fixate on A, and move from side to side by shifting your weight from one foot to the other. As you...
move to the right, the image of A is stationary on the retina, and the images of B and C move with you to the right, with B moving more slowly than C. Repeat the exercise while fixating on C. The image of C becomes stationary, and the images of A and B move to the left, opposite to your direction of movement, with B moving more slowly than A. Simply by changing your fixation point from A to C, the stationary image of A moves, the moving image of C becomes stationary, and the image of B reverses its direction of movement. To observe the direction of apparent movement reversing around the fixation point, perform the exercise a third time while fixating on B. The image of B becomes stationary, and the images of A and C move in opposite directions to one another, with A moving to the left, opposite to your direction of movement, and C moving with you to the right.

How does our demonstration show that motion parallax is an effective distance cue? As an observer moves in a three-dimensional environment while fixating on some point, images of objects farther from fixation move more quickly across the retina than do images of objects closer to fixation. If the object is between the observer and fixation, the faster the object’s image moves, the closer the object is to the observer. If the object is beyond the fixation point, the faster the image moves, the farther the object is from the observer. It is not completely clear, then, to state that objects closer to the observer move faster than do more distant objects. The key to a clear explanation of both the speed and direction of the image’s movement is to refer to the location of the object relative to the fixation point.

For a discussion of motion parallax within the larger context of motion perspective, we recommend that instructors consult Gibson (1950) and Haber and Hershenson (1980). For a simpler treatment of the phenomenon, we offer our explanation and demonstration. In either case, we encourage instructors to review the coverage of this complex phenomenon in their textbooks to assure that it is correct, complete, and clearly stated.

References


Notes

1. We thank Kimberly Purdy, who contributed critical discussion throughout the project; our psychology colleagues for lending us their textbooks; and Sally Quinnell for her helpful comments on this article.
2. Send correspondence to Yancy B. McDougal, Department of Psychology, University of South Carolina Spartanburg, Spartanburg, SC 29303; e-mail: ymcdougal@gw.uscs.edu.
and authority issues by focusing on author qualifications. The second criterion emphasized objectivity issues by discussing the potential for bias, as well as the difference between fact and opinion. The third criterion emphasized the accuracy and validity of the information by stressing the importance of appropriate references and discussing the concept of peer review. To illustrate the application of these criteria I performed a search on a topic that produced links of various quality. I chose a few links and showed how to apply the evaluation criteria to those sites. I then assigned the students different topics and asked them to locate both a high- and low-quality Web site for that topic. Although I did not grade these assignments, I did provide feedback when necessary.

Near the 10th week of the course, I assigned the students to one of eight small groups (ranging from 4 to 6 members), each with their own psychological disorder. Each student had to find four appropriate Web sites; I did not allow members of the same group to use duplicate sites. After finding four sites, each group member wrote a four-page paper that addressed the following topics: (a) a general description of the disorder, (b) etiology, (c) treatment options, and (d) long-term prognosis. The students then distributed a copy of their paper to each member of their group and wrote an evaluative summary for each paper, providing feedback on the quality of the Web information, the content quality of the written paper, and the quality of the writing. I provided guidance to the peer evaluations by distributing a handout listing specific issues to consider for each of these topics. The students also provided a descriptor of excellent, good, average, fair, or poor for each of these topics.

Each group then had to teach the class about their disorder. I set aside 2 class days as group work days, and each group created an outline for their disorder. During the presentations (requiring 4 class days), every student had to contribute. As each group discussed its disorder, I projected an overhead transparency of the outline onto a screen, and the student presenters made additional notes. The students in the audience received a photocopied packet of the outlines and took notes. All group members shared the same grade for their presentation and received an individual grade for their papers.

**Project Evaluation**

I evaluated my project by administering a student survey in three sections of the course (n = 112). Students stated that they spent an average of 3.25 hr (SD = 2.83) finding appropriate resources. They also claimed they spent an average of 3.63 hr writing their paper (SD = 2.10) and 2.47 hr writing their evaluations (SD = 1.38). The students responded to the other questions using 5-point rating scales (see Table 1). The students’ responses revealed that they were pleased with the potential for bias, as well as the difference between fact and opinion. The third criterion emphasized the accuracy and validity of the information by stressing the importance of appropriate references and discussing the concept of peer review. To illustrate the application of these criteria I performed a search on a topic that produced links of various quality. I chose a few links and showed how to apply the evaluation criteria to those sites. I then assigned the students different topics and asked them to locate both a high- and low-quality Web site for that topic. Although I did not grade these assignments, I did provide feedback when necessary.

Near the 10th week of the course, I assigned the students to one of eight small groups (ranging from 4 to 6 members), each with their own psychological disorder. Each student had to find four appropriate Web sites; I did not allow members of the same group to use duplicate sites. After finding four sites, each group member wrote a four-page paper that addressed the following topics: (a) a general description of the disorder, (b) etiology, (c) treatment options, and (d) long-term prognosis. The students then distributed a copy of their paper to each member of their group and wrote an evaluative summary for each paper, providing feedback on the quality of the Web information, the content quality of the written paper, and the quality of the writing. I provided guidance to the peer evaluations by distributing a handout listing specific issues to consider for each of these topics. The students also provided a descriptor of excellent, good, average, fair, or poor for each of these topics.

Each group then had to teach the class about their disorder. I set aside 2 class days as group work days, and each group created an outline for their disorder. During the presentations (requiring 4 class days), every student had to contribute. As each group discussed its disorder, I projected an overhead transparency of the outline onto a screen, and the student presenters made additional notes. The students in the audience received a photocopied packet of the outlines and took notes. All group members shared the same grade for their presentation and received an individual grade for their papers.

**Table 1. Questions and Student Ratings of the Project**

<table>
<thead>
<tr>
<th>Questions</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Before you took this course, how much prior experience did you have searching for information on the World Wide Web? a</td>
<td>3.03</td>
<td>1.17</td>
</tr>
<tr>
<td>2. On the day your group worked together in class, how well did everyone participate and contribute to the final finished product (i.e., the outline)? b</td>
<td>3.83</td>
<td>0.88</td>
</tr>
<tr>
<td>3. How well did the individual members of your group interact with one another? c</td>
<td>4.39</td>
<td>0.82</td>
</tr>
<tr>
<td>4. Were you personally pleased with the final outcome (i.e., the outline) produced by your group? d</td>
<td>4.04</td>
<td>0.76</td>
</tr>
<tr>
<td>5. Were you personally pleased with the paper that you wrote yourself? e</td>
<td>3.86</td>
<td>0.84</td>
</tr>
<tr>
<td>6. How much did you enjoy this assignment, in comparison to other term paper or major project requirements you have had in other classes? f</td>
<td>3.50</td>
<td>0.86</td>
</tr>
<tr>
<td>7. How difficult was this assignment, in comparison to other term paper or major project sorts of requirements you have had in other classes? g</td>
<td>3.05</td>
<td>0.85</td>
</tr>
<tr>
<td>8. How much do you feel you learned about your topic, in comparison to other term paper or major project sorts of requirements you have had in other classes? h</td>
<td>4.17</td>
<td>0.74</td>
</tr>
</tbody>
</table>

a Based on a scale ranging from 1 (none) to 5 (extensive). b Based on a scale ranging from 1 (little interaction) to 5 (much interaction; everyone participated equally). c Based on a scale ranging from 1 (not well) to 5 (very well). d Based on a scale ranging from 1 (not at all) to 5 (very much). e Based on a scale ranging from 1 (easy) to 5 (difficult). f Based on a scale ranging from 1 (little) to 5 (much).
Psychical thinking skills (Topping, 1998). Although my data do not show that the students actually learned new evaluation skills, the survey responses indicate that the students perceived that they learned new skills. Finally, the students got the opportunity to work in groups, an important skill in today's society (Uchida, Cetron, & McKenzie, 1996). Interestingly, many students' initial reaction to the project was that they would prefer to do it on their own. Many saw little value in group work and worried that the group grade would lower their own grade. The survey responses, however, indicated that by the end of the project, most saw its value.

I was also very pleased with the students' performance on the peer evaluations. Although many students spontaneously commented on the project survey that they disliked evaluating their fellow students, the quality of the evaluations was actually quite high. Most of the students were successful at distinguishing high- from low-quality sources. Many of them were also able to identify papers that were missing crucial information.

Concerns

Although grading the papers is time-consuming, they are interesting to read. A more pressing concern is to ensure that all group members contribute during all phases of the group project. A final concern is that the project requires seven classes. Although seven may seem too many, the approach is fairly efficient. For instance, the day devoted to Web evaluation training produces perceived benefits that extend well beyond this course. Additionally, because the students address effective therapies when they discuss their disorders, the presentations cover both the chapter on psychological disorders and much of the chapter on therapy. The groups discuss all of the major therapy approaches (i.e., biomedical, insight, cognitive, and behavioral) by the end of their presentations. Although sacrificing in-depth coverage, this technique does expose the students to the most common treatment methods.

Conclusions

My project emphasizes using Web site evaluation techniques and peer review to hone critical thinking skills. The survey responses show that the students worked hard, enjoyed the project, thought they learned quite a bit, and believed they developed useful real-world skills. Although I designed the project around the psychological disorders chapter, it could be adapted to almost any chapter given the availability of good Web information on many topics. The only requirement for successfully implementing the project is that instructors must have a good working knowledge of the Web, including how to find and evaluate information. Instructors lacking this knowledge can quickly familiarize themselves with the important issues by visiting one of the Internet evaluation sites listed on my home page.

References


Notes

1. Portions of this project were reported at the National Institute on the Teaching of Psychology conference in St. Petersburg Beach, FL, January 1999, and at the American Psychological Society conference in Miami, FL, June 2000.

2. Copies of all of the handouts and assignments used in the project are available at http://www.yk.psu.edu/~mac13/WebProject.htm. Additional links to other Web sites dealing with evaluation issues are available on my home page, http://www.yk.psu.edu/~mac13/index.htm.

3. Send correspondence to Mark A. Casteel, Penn State York, 1031 Edgecomb Avenue, York, PA 17403; e-mail: mac13@psu.edu.

Service Learning and Problem-Based Learning in a Conflict Resolution Class

Carole V. Wells

*Kutztown University of Pennsylvania*

This article involves a critical evaluation and discussion of the effectiveness of incorporating service learning and problem-based learning in a conflict resolution class. These strategies were effective pedagogical tools for introducing and clarifying patterns of conflict and conflict resolution.

Successful and productive resolution of conflict requires training, experience, and communication. Both service learning and problem-based learning (PBL) exercises provide integrative, experiential strategies that offer students an op-
Students who were unable to participate in community service were read, summarize, and critique a book related to conflict and conflict management.

Thirty-eight students engaged in service learning at 13 different agencies (see Table 1), and 10 students critiqued a book. Students kept a reflection journal on their service-based activities and submitted an experiential paper at the end of the semester. They not only described their service learning experiences but also applied a critical incident technique to an event they either observed or in which they participated. The event could be a decision made, a conflict, a change, or a problem resolved. They described the event and identified those individuals involved, why the issue was important, and who was influential in the outcome. Then, students detailed how the event was resolved or concluded, including the communication and conflict processes involved.

### PBL Exercises

Early in the semester, students discussed their previous positive and negative experiences working in groups, with a particular emphasis on effective group strategies. Class time also included mini-lectures on the benefits and potential pitfalls of group work, the group problem-solving process, the dynamics of PBL, and conflict styles and tactics. Students formed groups (i.e., six groups of 5 each and three groups of 6 each) during the second week of class and maintained the same group membership throughout the semester. Each group developed ground rules for behavior. Most group rules included being prepared for group work, respecting each other’s views and contributions, and attending all classes. Each group also agreed to designate rotating role responsibilities to a reporter, recorder, and discussion coordinator.

Students used an adaptation of a four-stage problem-based process proposed by several PBL researchers (see Allen et al.,

<table>
<thead>
<tr>
<th>Table 1. Service-Learning Experiences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organization</strong></td>
</tr>
<tr>
<td>1. Bible school</td>
</tr>
<tr>
<td>2. Child care center</td>
</tr>
<tr>
<td>3. County prison</td>
</tr>
<tr>
<td>4. Emergency shelter</td>
</tr>
<tr>
<td>5. Food service programs</td>
</tr>
<tr>
<td>6. Homeless shelter</td>
</tr>
<tr>
<td>7. Juvenile detention center</td>
</tr>
<tr>
<td>8. Meals on Wheels program</td>
</tr>
<tr>
<td>9. Nursing/retirement home</td>
</tr>
<tr>
<td>10. Salvation Army family center</td>
</tr>
<tr>
<td>11. Senior citizens council</td>
</tr>
<tr>
<td>12. Toys for Tots</td>
</tr>
<tr>
<td>13. University volunteer services</td>
</tr>
</tbody>
</table>

**Note.** N = 38.
students addressed different questions: Stage 1, What is the problem and why is it a problem?; Stage 2, What types of information are required?; Stage 3, What potential conflict styles could be used to resolve this controversy?; and Stage 4, What conflict resolution style serves the best interests of all concerned parties? Each group developed an understanding of arbitration, accommodation, avoidance, collaboration, competition, and compromise (Ruble & Thomas, 1976) as potential methods of conflict resolution.

The groups explored two instructor-initiated problem-based exercises, one focusing on the introduction of a school breakfast program and the other on exercising eminent domain. Students (a) engaged in detailed dialogue to explore the various facets of these two problems, (b) discussed what was needed to solve the problems and the necessary questions to be researched and answered, (c) identified personal strengths and weaknesses of group interactions, and (d) discussed conflict strategies to solve the particular problems.

Unlike the first two exercises, the third exercise required that students work in their groups, select a real-world conflict, research competing views about this problem, and propose a resolution method. After an initial session brainstorming ideas about potential conflict and conflict resolution issues, each group developed an action plan involving appropriate methods of information collection and individual assignments. Groups orally presented material related to each stage to the class with accompanying written discussion and conclusions.

Final student grades depended on the individual student’s performance on tests covering course content (35%), the service learning project paper (20%), and the three PBL group exercises (45%). Grades on the service learning and PBL exercises included group participation and ability to incorporate class material into the requirements.

Students’ Evaluations of Experiences

Students involved in service learning wrote papers describing and evaluating their experiences and integrating this assignment with what they had learned about conflict and conflict management. One student said, “I felt I was a mentor and used collaboration to bring even 2- or 3-year-old children to work together.” Another student remarked, “I learned that even though people may disagree with other’s opinions, they can only be resolved by understanding … . Taking the other person’s perspective into account is very helpful.” Several students reported feeling intrapersonal conflict. For example, “the main conflict that I felt in dealing with my prison visits was that I could not do enough.”

Feedback from the first PBL exercise on the development of a school breakfast program revealed that some groups thought (a) communication within the group was not well organized, (b) decisions were reached too quickly, (c) some group members were not heard and understood, and (d) group members did not make an effort to get to know each other and value each other’s points of view. Three students remarked that the subject of school breakfast was meaningless to them and, therefore, difficult to address. On the other hand, on a scale from 1 (very effective method) to 5 (very ineffective method), the mean student rating for this PBL exercise was 4.12 (SD = .73). In addition, in response to the question, “Overall, how effectively did your group work together?” the mean, on a scale from 1 (not at all effective) to 5 (very effective), was 4.23 (SD = .61).

Similar to the first PBL exercise, for the second PBL exercise (i.e., eminent domain) each student evaluated his or her group and individual member performances. As with their first experience, average group effectiveness was 4.72 (SD = .50), where 1 = not at all effective and 5 = very effective, and average effectiveness of PBL for understanding conflict resolution strategies was 4.57 (SD = .56), where 1 = not an effective method and 5 = very effective method.

Students reported they learned that gathering information is easier if the task is shared and that, although a discussion may initially seem controversial, a compromise can be reached and that “a group brainstorming can sometimes come up with good ideas quicker than an individual.” In response to suggestions for practical changes the group could make, students suggested (a) spending more time outside the class gathering information, (b) learning to work together better, (c) “giving everyone opportunities to speak without interruption,” (d) taking turns recording and note taking, (e) “talking more, because sometimes there was dead air,” (f) encouraging some group members not to do everything themselves, (g) expecting group members to pay more attention to detail, (h) “actively listening to make the group run smoother,” and (i) “giving specific tasks to prepare for each group meeting.”

For the third PBL exercise, groups selected conflicts such as school violence, hazing, college drinking, and capital punishment (see Table 2) and developed strategies and resolutions that would best serve the interests of all concerned parties. Four groups selected collaboration as the best strategy, three groups selected compromise, and one group (in the use of capital punishment) decided on review of the facts. Similar to their ratings on the second PBL exercise, student evaluations of the third exercise showed overall group effectiveness (M =

<table>
<thead>
<tr>
<th>Table 2. Group-Initiated Problem-Based Exercises</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Title IX and Equality in athletics</td>
</tr>
<tr>
<td>Male sports are being dropped to provide equity for women in athletics. Title IX is hurting male sports instead of helping female sports.</td>
</tr>
<tr>
<td>2. Megan’s law</td>
</tr>
<tr>
<td>There are convicted sex offenders in the community that people have a right to know about. Potential conflicts between community and lawmakers, judges releasing offenders into the community, and sex offenders.</td>
</tr>
<tr>
<td>3. Capital punishment</td>
</tr>
<tr>
<td>Conflict between those supporting and those against the death penalty. Conflicts among attorneys, state and federal governments, prisoners, and citizens.</td>
</tr>
<tr>
<td>4. Hazing</td>
</tr>
<tr>
<td>What constitutes hazing and what is against the law? It is a problem because of physical and emotional damages.</td>
</tr>
<tr>
<td>5. Reinstatement of Halloween parade in the borough</td>
</tr>
<tr>
<td>Traditional Halloween parade was canceled. Conflict among those in favor of reinstating the parade and those against the parade. Students versus Jaycees versus township versus police.</td>
</tr>
</tbody>
</table>

(continued)
6. Institution of measures to reduce school violence
   School wishes to add security cameras and guards, metal detectors, random locker searches, school uniforms, and video cameras. Conflicts among students, administrators, parents, school board, and larger community. Students view these measures as a violation of privacy and personal freedom. Also, suggest these measures are time consuming, irritating, and not conducive to learning. Both students and parents express concern that these measures will not get to the root of the problem.

7. College drinking—Stricter local drinking regulations
   Campus drinking among college students leads to vandalism/violence. Not enough activities in a small town.

8. Kosovo conflict
   Can Serbians and ethnic Albanians live together peacefully after the war? Problems because there are different religions, lack of job opportunity, Serbians want control over cultural resources, and Albanians want autonomy. 4.80, SD = .52, where 1 = not at all effective and 5 = very effective and the use of PBL as effective for understanding conflict resolution strategies (M = 4.58, SD = .54, where 1 = not an effective method and 5 = very effective method).

Conclusions

Both service learning and PBL exercises offer effective pedagogical strategies to enhance student development through active learning and participation, engagement in the community, critical thinking, and information on real-life conflictual situations. The blending of service and problem-based learning in this course provided active engagement, participation in the educational process, numerous community connections, and effective learning of conflict resolution strategies. This class of 49 “students,” including the professor, initiated a provocative, interactive, and reciprocal learning experience.

References


Notes

1. Portions of this article were presented at the State System of Higher Education annual conference on Advancing Teaching and Learning, Harrisburg, PA, March 2001.

2. I thank Randolph A. Smith and three anonymous reviewers for their constructive criticisms on earlier drafts of this article. I also thank R. Lorraine Bernotsky, Christopher Couch, and Elba Rohema for their many helpful suggestions and comments on the PBL exercises.

3. Send correspondence and requests for copies of the class materials and examples (e.g., the two instructor-initiated PBL exercises) to Carole V. Wells, Department of Psychology, Kutztown University, Kutztown, PA, 19530; e-mail: wells@kutztown.edu.

The Complexity of Student Responses to In-Class Debates in a Human Sexuality Course

Laura L. Finken
Creighton University

This study provides empirical evidence that in-class debates influence students’ attitudes about the debate topics. However, the impact of the debates on students was complex and varied by magnitude, topic, time, and even by students’ initial beliefs. Overall, the debates had a moderating impact on students’ attitudes, resulting in less extreme beliefs following the debates.

In-class debates hold appeal as a pedagogical technique because they have the potential to actively involve students, improve communication skills, force students to critically examine empirical evidence from two sides of an issue, and because students evaluate them positively (Elliot, 1993; Moeller, 1985; Smith, 1990; Waller, 1994). Indeed, the debate format is ideal for addressing controversies that relate to course content. Budesheim and Lundquist (1999) measured students’ attitudes at the beginning of the class and immediately following the debates. They found that the debates did
impact the students’ attitudes (sometimes strengthening positions, sometimes weakening positions) depending on the students’ initial beliefs and their assigned debate position.

However, a void in the literature exists given that no one has yet compared the immediate impact of in-class debates on students’ attitudes to the potential extended impact on students’ attitudes. Research on attitudes in general clearly indicates that people’s positions on issues fluctuate over time (e.g., Lowenthal & Lowenstein, 2001). Thus, based on Budesheim and Lundquist (1999), the expectation existed that students’ initial beliefs about the debate issues would influence their response to the debated material. Moreover, this study assessed attitude change both immediately following the in-class debate and at the end of the semester to identify potential differences between immediate and delayed effects.

Method

Participants

Participants included 151 undergraduate students enrolled in a human sexuality course across four consecutive semesters. The sample consisted of 33% men and 67% women and was predominantly (65%) European American. The participants ranged in age from 19 to 26 years with a mean age of 21.1 years.

Materials and Procedure

At the end of the first day in class, students answered questions about the debate topics as part of a larger pretest questionnaire: (a) Is pornography harmful to women?, (b) Should schools distribute condoms?, and (c) Should surrogate motherhood be outlawed? These questions were rated on a Likert scale ranging from 1 (no, absolutely not) to 10 (yes, absolutely). During the first week of class, the instructor assigned students to debate teams. Each team randomly selected their debate statement and position on this statement (e.g., Should schools distribute condoms?—Affirmative). The debates occurred at regular intervals throughout the semester. Prior to the debates, all students completed assigned readings on the issue and the involved debate teams did extensive research on the topic. Each debater had a designated role that necessitated his or her active participation during the debate. During debates that did not directly involve them, students observed as audience members and then had the opportunity to participate in the debate statement and position on this statement (e.g., Should schools distribute condoms?—Affirmative). The debates occurred at regular intervals throughout the semester. Prior to the debates, all students completed assigned readings on the issue and the involved debate teams did extensive research on the topic. During debates that did not directly involve them, students observed as audience members and then had the opportunity to participate in the debates for the last 15 min of the class. Following each debate, all students completed a reflective assignment about the topic and answered the same pretest questions about the debate topic. At the end of the semester, each student again answered the same questions about all of the debate topics. Finally, as part of the course evaluations at the end of the semester, students provided anonymous feedback about the debates.

Results

A repeated measures ANOVA was run for each debate topic with the attitude measure as the dependent variable, time as the within-subjects independent variable, and students’ initial position as the between-subjects independent variable. To create the initial position variable, the students were dichotomously divided for each topic by their pretest attitude measure into those who supported the debate statement (i.e., selected 6 to 10) and those who opposed the debate statement (i.e., selected 1 to 5). Table 1 presents the sample sizes and attitude means of these analyses; the variation in sample sizes resulted from alternative debate topics during various semesters and incomplete data from some students. As expected, all three debate analyses produced a main effect for initial position and a main effect for time. Moreover, analyses also yielded a significant two-way interaction between initial position and time for each of the debated issues. Follow-up analyses investigated the nature of these interactions for each of the topics using a Bonferroni-corrected significance level of $p < .0028$.

**Table 1. Sample Sizes and Mean Attitude Scores As a Function of Initial Position and Time**

<table>
<thead>
<tr>
<th>Debate Issue and Initial Position</th>
<th>Assessment Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Pretest</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
</tr>
<tr>
<td>Is pornography harmful to women?</td>
<td></td>
</tr>
<tr>
<td>Pornography is harmful</td>
<td>28</td>
</tr>
<tr>
<td>Pornography is not harmful</td>
<td>77</td>
</tr>
<tr>
<td>Should schools distribute condoms?</td>
<td></td>
</tr>
<tr>
<td>Schools should distribute condoms</td>
<td>21</td>
</tr>
<tr>
<td>Schools should not distribute</td>
<td>42</td>
</tr>
<tr>
<td>Should surrogate motherhood be outlawed?</td>
<td></td>
</tr>
<tr>
<td>Outlaw surrogate motherhood</td>
<td>17</td>
</tr>
<tr>
<td>Allow surrogate motherhood</td>
<td>27</td>
</tr>
</tbody>
</table>

Note. Different superscripts indicate significant differences between the assessment times at the $p < .0028$ level. For each issue, students rated their position on a Likert scale ranging from 1 (no, absolutely not) to 10 (yes, absolutely).
Condom Distribution Attitude Analysis

Follow-up analyses revealed that those students who initially believed that schools should distribute condoms did not experience a significant change in their attitude either immediately following the debate or by the end of the semester. Students who initially believed that schools should not distribute condoms experienced a dramatic change in their attitude immediately following the debate such that they then believed that schools should distribute condoms. However, by the end of the semester, their attitude toward condom distribution had shifted to a more neutral stance. Their end-of-the-semester view about condom distribution returned to the direction of their original position, but was still significantly tempered.

Surrogate Motherhood Attitude Analysis

Follow-up analyses revealed that students who initially believed that surrogate motherhood should be outlawed experienced a dramatic change in their attitude immediately following the debate such that they then believed that surrogate motherhood should be allowed. However, between the time of the debate and the end of the semester, they had gravitated back to their original belief so that there was no significant difference between their pretest attitude and final posttest attitude. The students who initially believed that surrogate motherhood should be allowed did not experience a significant change in their attitude immediately following the debate nor by the end of the semester.

Students’ Evaluation of Debates

To gain insight into the students’ experience of the debates, students anonymously rated the usefulness, level of difficulty, and their enjoyment of the debates on a 10-point scale ranging from 1 (not enjoyable) to 10 (very enjoyable) on the course evaluations at the end of the semester (N = 151). Overall, the students’ evaluation of the debates was favorable: they perceived debates to be useful (M = 8.15, SD = 2.00), relatively challenging (M = 7.36, SD = 1.58), and even somewhat enjoyable (M = 7.25, SD = 2.37).

Discussion

The results of this study indicate that in-class debates offer great promise to teachers as an effective approach to addressing controversial, course-related issues. Indeed, in this study in-class debates impacted students’ attitudes about issues related to human sexuality. However, the nature of that impact was complex—varying by magnitude, topic, time, and even by students’ initial beliefs. Because of the controversy surrounding issues related to sexuality in society and because of the complex nature of attitudes, it should not be surprising that the results of this study are complicated.

Most notably, the debates typically had a moderating effect on students’ attitudes immediately following the debates and even in several cases persisting through the end of the semester. This finding suggests that the in-class debates forced students to consider evidence from both sides of the issues. The purpose of the debates was not necessarily to change students’ opinions, but rather to promote thinking about issues from alternative perspectives. Indeed, it is reassuring to note that the debates typically only softened students’ original attitudes rather than leading to radical fluctuations in students’ underlying philosophical belief systems. Ideally, debates allow students to apply course material to the world in which they live, critically examine the evidence related to controversial issues, and reflect on their beliefs.

References


Notes

1. I thank Thomas Lee Budesheim for his comments on this article and for his support.
2. Send correspondence to Laura L. Finken, Department of Psychology, 2500 California Plaza, Creighton University, Omaha, NE 68178; e-mail: lfinken@creighton.edu.
An Evaluation of Industrial/Organizational Psychology Teaching Modules for Use in Introductory Psychology

Douglas C. Maynard
State University of New York at New Paltz

Peter D. Bachiochi
Eastern Connecticut State University

Ana C. Luna
State University of New York at New Paltz

Industrial/organizational (I/O) psychology has typically been neglected in introductory psychology textbooks and courses. The Society for Industrial and Organizational Psychology (SIOP) recently developed a series of teaching modules for introducing I/O psychology. We evaluated 4 modules with 12 samples \( N = 333 \) for student learning and intentions and for student and instructor reactions. Overall, student knowledge of I/O concepts increased after presentation of the modules. Additionally, students were more likely to want to take a course in I/O psychology after the presentation. Finally, students found the presentations interesting and easy to understand. All instructors reported that the modules made presenting a lecture on I/O easy and that they were considering the use of the modules for future sections.

Instructors of introductory psychology courses have the difficult but important task of choosing a limited set of topics to cover. The material in an introductory psychology course likely affects student decisions to (a) declare psychology as a major or minor, (b) take other psychology courses, and possibly even (c) pursue graduate training in a given area in psychology. In addition, the mix of topics chosen by the instructor shapes student perceptions of the field of psychology. If this mix is not representative, students may finish the course with a distorted impression of the discipline.

Because of the diversity of psychology, it is inevitable that some areas will not receive appropriate attention in introductory psychology courses. For instance, industrial/organizational (I/O) psychology, the study of human thought and behavior in the workplace, is typically neglected. Introductory psychology textbooks devote little space to I/O psychology (Carlson & Millard, 1984; Griggs, Jackson, Christopher, & Marek, 1999). Miller and Gentile (1998) found that, of 25 introductory psychology textbook topics, instructors assigned I/O psychology least often. Given that textbooks normally do not include a chapter or section on I/O psychology and most instructors are not I/O psychologists themselves, it is not surprising that the topic is frequently overlooked.

There are several reasons why I/O psychology should be part of the introductory psychology experience. First, more psychology departments are offering I/O courses (Perlman & McCann, 1999); any exposure to the field would help students make more informed decisions about taking an I/O course. Second, discussions of I/O psychology can be used to demonstrate how core psychological concepts (e.g., motivation) can be applied to real-world problems. Third, recent data have suggested that psychology undergraduates are more likely to find employment in business and management than in any other occupational area (“A Look at Recent Baccalaureates in Psychology,” 2000). Finally, the general public seems to have a lack of awareness of the field of I/O psychology and the work that I/O psychologists do (Gasser et al., 1998).

The SIOP Instructor’s Guide

To aid instructors who wish to include I/O psychology in introductory psychology, the Education and Training Committee of the Society for Industrial/Organizational Psychology (SIOP; Division 14 of the American Psychological Association), created An Instructor’s Guide for Introducing Industrial and Organizational Psychology (Bachiochi et al., 1999). The committee designed the guide, organized as a series of six distinct modules, for instructors with relatively little exposure to the field. The modules (Evaluating Work Performance, Leader-Member Interactions, Leadership and Gender Stereotypes, Motivation and Performance, Sexual Harassment, and Workplace Diversity) are available in PowerPoint® format on SIOP’s Web site, along with suggestions for exercises, discussions, supplemental readings, and videotapes (see the Reference section for the Web address).

The first module, a brief overview of I/O psychology, is a lead-in to any of the six specific topic modules; together, they can be presented in either a 50- or 75-min class period. The committee chose topics that would (a) generate real student interest and (b) complement other popular introductory psychology topics, such as social psychology and motivation (Bachiochi & Major, 1999). We present an outline of the sexual harassment module in Table 1 as an example of the
structure used in many of the modules (i.e., lesson objectives, background information/definitions, current issues, an exercise/discussion, and a conclusion indicating the role of I/O psychologists with the topic).

In this study, we evaluated four modules in this guide (i.e., the Overview, Leadership and Gender Stereotypes, Sexual Harassment, and Workplace Diversity modules). We examined student and instructor reactions to the modules as well as changes in student knowledge and intentions (e.g., intention to take a course in I/O psychology). We did not evaluate all modules because we gave instructors the freedom to choose the module topic they found most interesting or believed would fit well with the other topics discussed in their course. We gave instructors this choice because it more closely reflects how the instructor’s guide would actually be used, as opposed to having us choose a module for the instructor.

Method

Participants

Four instructors presented modules as lectures to students (N = 333) in 10 different introductory psychology courses at two medium-sized Northeastern public universities (Samples 1 to 10). Table 2 presents sample and instructor information.

Comparison groups. Sample 11 had the same instructor of record as Sample 1 and met on the same day, but did not receive an I/O psychology presentation. We collected knowledge and intentions data for both samples. Sample 12 consists of students from the classes in Samples 3 through 10 who were not in attendance on the day of the I/O psychology presentation, but were in attendance on the subsequent class to complete the posttest measures. Samples 11 and 12 represent comparison groups that we used to identify whether any pre–post differences might be due to testing effects (e.g., practice or sensitization). Although there may be motivational differences between students in Sample 12 and students in Samples 3 through 10 due to self-selection, we believe that the availability of posttest data for nonattendees represented an additional (although imperfect) opportunity to test the effectiveness of the modules.

The majority of the participants were freshmen (52%) or sophomores (38%). A total of 26% of the participants had not yet declared a major. The most common majors for participants who had declared were elementary education (17%), psychology (15%), and business (15%).

Measures

Student knowledge. We wrote two multiple-choice questions for each module to assess knowledge of content.
Thus, students in Samples 1 and 11 completed only the two overview module questions; students in the remaining samples answered four multiple choice questions (two each for the overview and topic modules).

**Student intentions.** We asked students to provide their intentions toward (a) taking an I/O psychology course, (b) learning more about I/O in the future, (c) declaring psychology as a major, and (d) pursuing employment in I/O related areas (e.g., human resources). We measured these intentions with a scale ranging from 1 (strongly disagree) to 5 (strongly agree).

**Student reactions.** We assessed student reactions to the I/O modules with seven items (e.g., “The I/O Psychology lecture was intellectually stimulating.”) using a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The average coefficient alpha for the student reactions scale was .84.

**Instructor reactions and intentions.** We assessed instructor reactions to the teaching modules with seven items (e.g., “The teaching module was easy to present to the students”). The first and second authors (Instructors 2 and 3) did not complete the instructor reactions questionnaire. In four samples, guest lecturers presented the modules rather than the instructor of record (see Table 2). In these cases, both provided their reactions by answering items that were relevant to them and answering not applicable to the remaining items. The first and second authors (Instructors 2 and 3) did not complete the instructor reactions questionnaire. The remaining four instructors who completed the reactions questionnaire included one guest lecturer (Instructor 1), one instructor of record who presented the module (Instructor 4), and two instructors of record who did not present the module themselves (in Samples 1 and 2).

### Procedure

At the beginning of class, students completed the pretest questionnaire, which included the student knowledge and intentions measures. The instructor then presented the modules. For Sample 1, the instructor gave the pretest and presented the overview module during the last 20 min of the class period. As comparison groups, Samples 11 and 12 did not receive the pretest, module presentation, or student reactions measure. All classes were 75 min in length except for Samples 3 and 4, which had 50-min classes.

Two days after the presentation of the modules, instructors administered the posttest questionnaire, which included the student knowledge, reactions, and intentions measures. Note that because students received the posttest during the very next class meeting, there is little chance that nonattendees

### Results

#### Change in Student Knowledge

**Preclass versus postclass knowledge.** We assessed student learning in two ways. First, we conducted paired-samples t tests to compare the class-specific knowledge before and after the I/O presentation for students who were present. As shown in Table 3, student knowledge after the presentation was significantly higher than student knowledge before the presentation for all but the Workplace Diversity module.

**Knowledge of attendees versus nonattendees after the presentation.** Second, we conducted independent samples t tests to examine the postpresentation knowledge for students who received the I/O lecture versus those from relevant comparison groups who did not. As shown in Table 4, students who received the I/O lecture answered significantly more questions correctly than those who did not receive the lecture, with the exception of the Workplace Diversity questions.

**Student Intentions**

We conducted paired-samples t tests to assess changes among attendees in various intentions related to I/O psychol-

---

**Table 3. Change in Student Knowledge Across Samples**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Prepresentation</th>
<th>Postpresentation</th>
<th>df</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>General I/O knowledge</td>
<td>0.94 (.73)</td>
<td>1.45 (.64)</td>
<td>214</td>
<td>4.66*</td>
</tr>
<tr>
<td>Leadership knowledge</td>
<td>1.50 (.65)</td>
<td>1.81 (.40)</td>
<td>25</td>
<td>2.13*</td>
</tr>
<tr>
<td>Sexual harassment knowledge</td>
<td>0.85 (.73)</td>
<td>1.24 (.77)</td>
<td>134</td>
<td>5.90**</td>
</tr>
<tr>
<td>Workplace diversity</td>
<td>1.26 (.82)</td>
<td>1.55 (.68)</td>
<td>30</td>
<td>1.87</td>
</tr>
</tbody>
</table>

*Note.* I/O = Industrial/organizational.

*a* Sample 2. *b* Samples 3 through 8. *c* Samples 9 and 10. *p < .05. **p < .01.
ogy. Students were significantly more interested in taking an I/O course following the presentation than prior to the presentation (see Table 5). There were no other significant changes in student intentions across all samples.

**Student Reactions**

Reactions to the I/O presentation were very positive overall. Students enjoyed the presentation (75.1%), found it interesting (77.3%), and found it intellectually stimulating (59.3%). Eighty-two percent believed the presentation broadened their knowledge of the different areas of psychology and, perhaps most important, 66% agreed that I/O psychology should be included in future general psychology courses. Finally, 88.9% believed the presentation was organized and made sense, and 90% believed it was easy to understand.

**Instructor Reactions and Intentions**

All responding instructors agreed or strongly agreed that the teaching module made the presentation easy to prepare and present, the content was written at an appropriate level for introductory students, the notes that supplemented the slides were helpful, and the slides themselves were easy to understand. Instructors disagreed on whether the modules stimulated good class discussion and that the student response was enthusiastic. In sum, instructors were positive on five reaction items and mixed on two items. In terms of intentions, all instructors agreed that they would consider use of an I/O teaching module in future sections of introductory psychology, and most (75%) planned to cover I/O in future sections.

**Discussion**

For all samples, both students and instructors had positive overall reactions to the modules. With the exception of the Workplace Diversity module (Samples 9 and 10), student content knowledge increased from pretest to posttest, and students who attended the presentation outperformed nonattendees on the knowledge measures.

We found mixed results with regard to student intentions. Students were more likely to want to take an I/O psychology course after the presentation, although this effect was small. There were no changes in other student intentions. Although this finding was disappointing, it may be unrealistic to expect that a diverse sample of students would, as a whole, be more interested in I/O psychology after learning about it. Before any unfamiliar topic is presented, intentions for students are likely to be moderate, due to a lack of knowledge. After hearing the presentation, some students will likely want to learn more, but other students may find the topic uninteresting.

Several limitations to this study and directions for future research should be noted. First, the knowledge questions for two of the modules (Leadership and Gender Stereotypes and Workplace Diversity) were too easy; pretest and nonattendee performance were quite high. The resulting ceiling effect may be responsible for the nonsignificant changes in learning for the latter module. Assessing student knowledge with a larger number of more challenging items could prevent such an effect in future evaluations. Second, we evaluated only three of the six modules in the instructor’s guide; future research is needed to evaluate the remaining modules. Third, future investigations could include additional intention items, particularly for nonmajors, such as:

<table>
<thead>
<tr>
<th>Questions</th>
<th>Preclass</th>
<th>Postclass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Want to take an I/O course</td>
<td>2.96</td>
<td>3.08</td>
</tr>
<tr>
<td>May become psychology major</td>
<td>2.87</td>
<td>2.84</td>
</tr>
<tr>
<td>Want to learn more about I/O</td>
<td>3.26</td>
<td>3.26</td>
</tr>
<tr>
<td>May seek I/O related employment</td>
<td>2.81</td>
<td>2.82</td>
</tr>
</tbody>
</table>
as declaring a psychology minor or taking additional psychology courses relevant to their major.

Several instructors provided useful insights that might aid future instructors in the successful use of these modules. One instructor suggested that the modules will work best when instructors tailor the module to their own style of teaching (e.g., use of overheads), rather than presenting it exactly as packaged. Another observation was that, although most students have some work experience, most have not held full-time positions. Accordingly, instructors should orient any additional examples they use to the types of jobs commonly held by high school and college students.

In conclusion, the SIOP instructor’s guide appears to be a useful resource for introducing the topic of I/O psychology in introductory psychology courses. The use of this guide has several benefits. First, the modules expose students to an important area of applied psychology; this exposure gives them a more comprehensive view of the field. In addition, students will be better able to make an informed decision about whether to take an entire course in I/O psychology or consider a career in a related area, such as human resources. We hope that introductory psychology instructors, provided with these teaching materials to compensate for a lack of textbook coverage, will devote class time to this important and growing subdiscipline of psychology.

References


Notes

1. We thank Maryalice Citera, Margaret Letterman, David L. Morse, Jonathan D. Raskin, and Leo Schneiderman for their assistance in this research. We also thank Todd Thorsteinson and three anonymous reviewers for helpful comments on an earlier version of this article. Finally, we acknowledge the SIOP Education and Training Committee for their work in developing the instructor’s guide, which can be found at http://www.siop.org/Instruction/InGuide.htm.
2. Send correspondence and requests for copies of the entire set of measures to Douglas C. Maynard, Department of Psychology, State University of New York at New Paltz, 75 South Manheim Boulevard, Suite 6, New Paltz, NY 12477–2440; e-mail: maynardd@newpaltz.edu.
APPENDIX: CITATION INFORMATION

All articles in this book appeared originally in the journal, *Teaching of Psychology*. This appendix provides the year and volume number of original publication, plus page numbers, to facilitate proper citations of these articles.

Section I. Approaches and Issues in Teaching Introductory Psychology

1. Approaches to the Introductory Course
   - Kretchmar, 2001, 28, 5-10
   - Peck, Ali, Matchock, & Levine, 2006, 33, 167-170
   - Lawson, Bodle, & McDonough, 2007, 34, 90-93
   - Griggs & Jackson, 2007, 34, 181-186
   - Kowalski & Taylor, 2009, 36, 153-159
   - Elicker, Snell, & O’Malley, 2010, 37, 36-40
   - Hermann, Foster, & Hardin, 2010, 37, 79-84
   - Ritchey & Bott, 2010, 37, 262-266

2. Research Participation for Introductory Students
   - Koschmann & Wesp, 2001, 28, 105-108
   - Ferrari & McGowan, 2002, 29, 29-32
   - Berthold, Hakala, & Goff, 2003, 30, 55-58
   - Rosell, Beck, Luther, Goedert, Shore, & Anderson, 2005, 32, 95-99
   - Trafimow, Madson, & Gwizdowski, 2006, 33, 247-249
   - Thieman, Clary, Olson, Dauner, & Ring, 2009, 36, 160-168
   - Sacco & Bernstein, 2010, 37, 28-31

3. Active Learning
   - Butler, Phillmann, & Smart, 2001, 28, 257-259
   - Bleske-Rechek, 2001, 28, 260-262
   - Harton, Richardson, Barreras, Rockloff, & Latané, 2002, 29, 10-15
   - Katz, 2003, 30, 314-316
   - Muir & van der Linden, 2009, 36, 169-173

4. Examinations: Test Items
   - Scialfa, Legare, Wenger, & Dingley, 2001, 28, 11-15
   - Carroll, 2001, 28, 98-100
5. Examinations: Student Preparation
   - Balch, 2001, 28, 272-274
   - Perlman & McCann, 2002, 29, 51-53
   - Fleming, 2002, 29, 115-119
   - Gurung, 2005, 32, 239-241
   - Zamboanga, Padilla-Walker, Hardy, Thompson, & Wang, 2007, 34, 158-162
   - Balch, 2007, 34, 219-225

6. Examinations: Testing Factors
   - LoSchiavo & Shatz, 2002, 29, 138-140
   - Skinner, 2004, 31, 111-113
   - Dickson & Miller, 2006, 33, 39-40
   - Tal, Akers, & Hodge, 2008, 35, 26-28

7. Examinations: Course Design Factors
   - Leeming, 2002, 29, 210-212
   - Kouyoumdjian, 2004, 31, 110-111
   - Balch, 2005, 32, 29-34
   - Padilla-Walker, 2006, 33, 236-239
   - Landrum, 2007, 34, 177-180

8. Examinations: Effects of Study Guide and Pedagogical Aid Use
   - Gurung, 2003, 30, 92-95
   - Gurung, 2004, 31, 164-166
   - Dickson, Miller, & Devoley, 2005, 32, 34-39
   - Dickson, Devoley, & Miller, 2006, 33, 40-42

9. Students’ Course Preferences and Knowledge about Psychology
   - Gaultney & Cann, 2001, 28, 84-87
   - Thompson & Zamboanga, 2003, 30, 96-101
   - Landau & Bavaria, 2003, 30, 311-314
   - DiClementi & Handelsman, 2005, 32, 18-21

10. Introductory Textbooks: Selection and Student Use
     - Landrum & Hormel, 2002, 29, 245-248
     - Sikorski, Rich, Saville, Buskist, Drogan, & Davis, 2002, 29, 312-313
     - Griggs, Jackson, & Marek, 2002, 29, 314-316
11. Introductory Textbooks: Content
   Griggs & Marek, 2001, 28, 254-259
   Sheldon, 2002, 29, 281-285
   Steuer & Ham, 2008, 35, 160-168
   Haselhuhn & Clopton, 2008, 35, 205-209

12. Extra Credit
   Thorne, 2000, 27, 204-205
   Hardy, 2002, 29, 233-234
   Boniecki & Moore, 2003, 30, 224-227

Section II. Technology in Teaching Introductory Psychology

1. Online Delivery of the Course
   Waschull, 2001, 28, 143-147
   Newlin & Wang, 2002, 29, 325-330
   Waschull, 2005, 32, 190-192
   LoSchiavo & Shatz, 2005, 32, 246-248

2. Computer Assisted Instruction
   Schult & McIntosh, 2004, 31, 209-211
   Pettijohn & Perelli, 2005, 32, 54-56
   Hardin, 2007, 34, 53-57
   Elicker, O’Malley, & Williams, 2008, 35, 126-131
   Daniel & Woody, 2010, 37, 199-203

3. Posting Course Lecture Notes
   Vandehey, Marsh, & Diekhoff, 2005, 32, 49-52
   Cornelius & Owen-DeSchryver, 2008, 35, 6-12
   Hove & Corcoran, 2008, 35, 91-95

4. Electronic Review Tools
   Pemberton, Borrego, & Cohen, 2006, 33, 145-147
   Paul, Messina, & Hollis, 2006, 33, 276-279
5. **Electronic Student Response Systems**
   Poirer & Feldman, 2007, 34, 194-196
   Stowell & Nelson, 2007, 34, 253-258
   Cleary, 2008, 35, 42-44
   Morling, McAuliffe, Cohen, & DiLorenzo, 2008, 35, 45-50
   Stowell, Oldham, & Bennett, 2010, 37, 135-140

6. **Computerized Quizzing**
   Brothen & Wambach, 2001, 28, 292-294
   Brothen & Wambach, 2004, 31, 62-64
   Grimstad & Grabe, 2004, 31, 143-146

### Section III. Demonstrations and Activities in Introductory Psychology

1. **General**
   Merwin, 2003, 30, 242-244
   Balch, 2006, 33, 249-252

2. **Research Methods and Statistics**
   Griggs, 2000, 27, 132-135
   Lawson, 2003, 30, 251-253
   Wesp & Eshun, 2005, 32, 234-236
   Balch, 2006, 33, 132-134
   Treadwell, 2008, 35, 184-188

3. **Sensation and Perception**
   Haws & Oppy, 2002, 29, 147-150
   Prull & Banks, 2005, 32, 103-106

4. **Learning**
   Batsell, 2006, 33, 129-131

5. **Memory**
   Hoyert & O’Dell, 2000, 27, 272-273
   Clump, 2006, 33, 51-53

6. **Cognition**
   Swinkels, 2003, 30, 120-122
Griggs, 2003, 30, 143-145

7. Social Psychology
   Jackson, 2000, 27, 273-276
   Snyder, 2003, 30, 59-61
   Christopher, Walter, Marek, & Koenig, 2004, 31, 199-202
   Hebl & King, 2004, 31, 260-262

8. Personality
   Lipsitz & Gifford, 2003, 30, 58-59
   Nier, 2004, 31, 53-54
   Miserandino, 2006, 33, 265-268
   Miserandino, 2007, 34, 37-40

9. Abnormal Psychology
   Casteel, 2003, 30, 258-260

10. Industrial/Organizational Psychology
    Maynard, Bachiochi, & Luna, 2002, 29, 39-43