Encouraging Active Learning Can Improve Students' Performance on Examinations

Janice D. Yoder and Catherine M. Hochevar
University of Akron

We tested the hypothesis that students in psychology of women classes would perform better on materials covered by multiple-choice exams when the first author presented these materials with active learning versus lecture, autonomous readings, and video presentations alone. Across 3 classes, we coded exam items according to how the instructor presented relevant materials and recorded classwide performance. Both between and within classes, students' performances were better on items testing materials covered with active learning techniques compared to other formats. These data provide empirical support for the efficacy of active learning techniques.

Active learning is a buzz phrase that captures the teaching technique promoted by learner-centered as opposed to content-centered instruction (Halonen, Brown-Anderson, & McKeachie, 2002). Active learning is "anything that involves students in doing things and thinking about the things they are doing" (Bonwell & Eison, 1991, p. 2). The Committee on Developments in the Science of Learning (Bransford, Brown, & Cocking, 2000) further refined this broad definition to emphasize the importance for students to control their learning, and Mayer (2004) highlighted the importance of structured, guided discovery. What makes active learning fundamentally active is the cognitive processing demanded to find patterns in materials provided, organize these patterns into meaningful clusters, understand under what conditions this knowledge is useful, and retrieve it fluently (Bransford et al., 2000). Active learning thus goes beyond the simple availability of information to facilitate students' self-discovery of knowledge (Mathie et al., 1993) and consequent performance (Butler, Phillmann, & Smart, 2001; Lake, 2000; Lonka & Ahola, 1995; Schwartz & Bransford, 1998).

McKeachie (2002) designated discussion as the prototypic method for active learning, although others have expanded this operationalization to include visually based instruction, writing in class, problem solving, computer-based instruction, cooperative learning, debates, drama, role playing, simulations, games, and peer teaching (Sivan, Leung, Woon, & Kember, 2000). *Teaching of Psychology* offers many examples of these techniques, and some articles provide empirical evidence of their effectiveness toward promoting students' learning (e.g., Butler et al., 2001). However, we know of no study to date that combined a variety of active learning techniques to explore their overall effectiveness relative to nonactive learning approaches. We designed this study to begin to fill this gap.

In this study, we compared students' performance on the same multiple-choice exam items across psychology of women classes taught in the spring semesters of 2001, 2002, and 2003. Across these three classes, the same instructor (the first author) using the same text moved toward devoting more classroom time to active learning and relied increasingly on students' autonomous textbook reading and peer learning to cover some materials. We hypothesized that both within and between classes, classes' exam performance would improve when the instructor presented course materials using active learning compared to other techniques. Given our progression toward more active learning across the 3-year period studied, we used the most recently taught class (2003) as the base, tracking changes from 2001 to 2003 as well as from 2002 to 2003.

Often instructors hesitate to move toward expanded inclusion of active learning approaches in their classes because classroom coverage of some course content is lost. Although active learning, as we broadly define it here, need not use much classroom time, it was clear across these classes that as active learning coverage increased, the instructor depended more heavily on autonomous textbook readings. She was concerned that this change in instructional approach might result in poorer performance on exam materials not covered in class. Therefore in 2003, the instructor formed in-class student teams that she encouraged to work together both inside and outside class time. This structure draws on peer teaching and learning, which depends on students to learn from other students who share their level of expertise (see Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Nelson, 1996). Given these supports, we predicted that students would perform equally well on identical items covered by lecture during the prior year but subsequently left unmentioned in class with coverage only through textbook readings. To test this hypothesis, we examined class performance on a subset of exam items that tested materials covered in lecture in classes prior to 2003 then only in readings in 2003.

Method

Participants

Participants were 45 students in 2001 (33 White women, 3 African American women, 1 other; 5 White men, 3 African American men), 37 in 2002 (26, 4, 0; 6, 1, respectively), and 38 in 2003 (30, 4, 2; 2, 0, respectively) enrolled in a 400-level undergraduate psychology of women class at a large, public university taught by the first author and requiring Yoder (1999, 2003) as the text. The expertise of the instructor arguably was stable across the 3 years as she had taught this course regularly for the past 11 years and had 24 years of teaching experience as of 2001. The majority of students were women (82% in 2001, 81% in 2002, and 95% in 2003), but the classes did not differ significantly in their proportions of women, $\chi^2(2, N = 120) = 3.65, p = .16$. The representation of African American students in each class (13%, 14%, and 16%, respectively) was consistent with the proportion across the university (15%). The mean grade point average (GPA) of the three classes did not differ ($M = 3.03$, $SD = .60$; $M = 3.18$, $SD = .53$; $M = 3.04$, $SD = .50$, respectively), $F(2, 104) = .79, p = .46$. The majority of students in each class majored in an area within arts and sciences (64%, 73%, and 76%, respectively), and they were upper level students (82%, 93%, and 84%, respectively).

Procedure

We obtained class GPAs without identifying individuals and secured institutional research board approval to use aggregated class data. We archived all data and coded retrospectively; at the time of instruction, there were no plans to conduct this study so that experimenter effects from the instructor–researcher were unlikely. To compare the 2001 class with 2003 and again the 2002 class with 2003, we created two 4 × 2 mixed designs, with four presentation patterns serving as a between-items variable and aggregated class performance on each item across the compared classes as a within-items, repeated measure. Note that the unit of analysis was exam items, not students. Thus the major variables of interest were (a) classwide performance (proportion of correct answers) on identical multiple-choice items used repeatedly across classes and (b) a preliminary coding by the instructor of presentation method for materials relevant to each exam question within each class (active learning, lecture, readings only, or video without discussion). To do this initial coding, the first author drew heavily on detailed records of class plans she maintained for each of the three classes as support materials for revisions of her textbook.

Content coverage using active learning techniques, exclusively or in addition to lecture, readings, and videos, ranged from brief to extended discussions of materials classwide or in small groups, exercises, simulations, and demonstrations as well as completion and discussion of relevant measurement scales (e.g., the Objectified Body Consciousness Scale; McKinley & Hyde, 1996). We coded materials as book only when these materials were not discussed in class. Video presentations involved the presentation of audiovisual materials without follow-up discussion (video presentations preceded or followed by active discussion as examples of active learning). Across all three classes, to facilitate students' preparations, students received study guides, which outlined the materials to be covered, in advance of exams.

This coding of presentation style was only the first step toward creating four levels of the independent variable of interest, that is, change in style across classes using the same exam items. Two experimental conditions involved exam items (a) for which the instructor changed from nonactive learning (combining lecture, book only, and video only) to having at least some active learning component for those materials in the subsequent class and (b) that drew on active learning in classes being compared. We also created two control conditions wherein materials (c) employed the same specific nonactive learning approach in both classes (e.g., lecture in both) or (d) used a different nonactive learning strategy across classes (e.g., lecture in the first class and readings only in the next class). These control conditions examined unchanged and changed presentation techniques, respectively, appropriately paralleling the two experimental conditions but without any active learning component. This design and its cell sizes appear in Table 1.

In 2001, the instructor used two 20-item exams, with 38 of these 40 items being repeated in the next 2 years. In both 2002 and 2003, students completed three 40-item exams, of which 105 of the 120 items were repeated.[1] Engagement in active learning increased across the three classes with 13% (5 of 39 items) of exam items in 2001 testing materials presented via active learning, 20% (22 of 108 items) in 2002, and 27% (29 of 108 items) in 2003. Commensurate with this shifting emphasis, reliance on lecture declined (77% or 30 items; 35% or 38 items; 19% or 21 items, respectively) and reliance on autonomous readings increased (7% or 3 items; 36% or 39 items; 48% or 52 items, respectively). Each year represented a significant shift toward greater reliance on active learning compared to the other presentation styles combined, 2001 to 2002: $\chi^2(1, N = 39) = 12.90, p < .001$, Cramer's $\phi = .58$; 2002 to 2003: $\chi^2(1, N = 108) = 49.82, p < .001$, Cramer's $\phi = .68$.

To check on the comparability of exam items used to test materials covered with active versus the other learning approaches, we further coded exam items using an updated revision of Bloom's taxonomy (Anderson & Krathwohl, 2001). Higher levels of learning (coded 1) were tapped by items that required application, analysis, evaluation, or synthesis in contrast to knowledge or rote memorization and basic understanding (coded 0). Across all three classes, the proportion of items classified as higher (68% to 69% of all items within each class) and lower level was the same for active and other learning items: 2001: $\chi^2(1, N = 39) = .31, p = .58$; 2002: $\chi^2(1, N = 108) = .001, p = .97$; 2003: $\chi^2(1, N = 108) = .004, p = .95$. These comparisons suggest that exam items were of similar difficulty across presentation styles and tapped similar levels of cognitive processing of information.

---

[1] Four items were excluded: From 2001 to 2003, one item shifted from active learning to lecture; from 2002 to 2003, one item changed from active learning to lecture and two items changed from active learning to book only. Other assessment techniques used included student presentations, short essay exam questions, and papers, but we did not analyze these. The use of multiple-choice questions creates a more stable test of the hypotheses. The focus of these additional assignments was parallel across classes so that their impact, if any, would be similar.

Results

To test for performance changes across 2001 to 2003, we used a mixed-design ANOVA that explored changes in aggregate class performance across the 2 years (the within-items variable) within each of the four item categories (between-item variable). The predicted interaction of performance changes across presentation patterns was significant, Hotelling's $F(3, 34) = 4.56$, $p = .009$, $\eta^2 = .29$. As shown in Table 1 (across rows), this interaction occurred due to improved performance within those items testing materials where presentations shifted toward active learning, $t(8) = 3.03$, $p = .016$, $r = .79$. In contrast, exam performance across these two classes remained stable within the other three presentation categories. Looking within classes (Table 1's subtotal columns), students in 2001 scored better on the 4 items taught with active learning techniques than on the other 34 items, $t(35.8) = 5.84$, $p < .001$, $\eta^2 = .11$, with less variance in scores for the active learning items, $F(1, 36) = 5.81$, $p = .02$. Students in 2003 performed similarly on active compared to nonactive learning items repeated from the 2001 class, $t(36) = 1.62$, $p = .11$.

Turning to comparisons of the 2002 with the 2003 class, again the expected interaction occurred, Hotelling's $F(3, 101) = 6.27$, $p < .001$, $\eta^2 = .16$ (see Table 1). The follow-up analyses replicated the same pattern of findings outlined for the 2001 and 2003 class comparisons. In contrast to across-class performance, which remained stable within consistently active learning or nonactive learning presentations of materials, students scored higher on exam items where we replaced nonactive learning presentations with active learning ones, $t(9) = 3.23$, $p = .01$, $r = .77$. Like the 2001 class, students in the 2002 class scored higher on items covering materials taught with active learning techniques than on those items taught using nonactive learning, $t(43.3) = 3.48$, $p = .001$, $\eta^2 = .06$, again with less variance among active learning items, $F(1, 103) = 6.49$, $p = .01$. Using the items repeated from 2002 to 2003, we found a subtotal difference between active and nonactive learning item performance in 2003 favoring the 29 active learning items, $t(79.1) = 2.71$, $p = .008$, $\eta^2 = .05$, again with lower variance among active learning items, $F(1, 103) = 7.39$, $p = .008$.

Finally, to explore our speculation that, with compensatory supports, lost lecture coverage need not impede exam performance, we compared performances between classes on items covered by lecture previously but presented in 2003 only through textbook readings. Considering 12 such items from 2001 ($M = .77$, $SD = .12$) to 2003 ($M = .77$, $SD = .12$), no differences appeared, $t(11) = .19$, $p = .86$, $r = .75$. The same was true on the different set of 13 items changed from lecture to book-only reading from 2002 ($M = .73$, $SD = .11$) to 2003 ($M = .71$, $SD = .18$), $t(12) = .35$, $p = .73$, $r = .65$.

Discussion

The patterns across the data analyses are clear and generally consistent: Both within a class and between classes, classes scored higher and less variably on items testing materials presented via active learning compared to lecture, autonomous readings, or video without discussion coverage. The one exception occurred with items from 2001 repeated in 2003 for which students' better performance on actively learned materials was in the expected direction but did not reach statistical significance. In contrast, classes of students with similar GPAs scored comparably on items when presentation techniques were similar, suggesting that differences in students did not account for differences in test performance. Between-class comparisons kept the items and the materials covered constant, reducing the plausibility of explanations involving the variable difficulty of materials and items (also supported by analyses showing item comparability across Bloom's levels of learning). The patterns of data were repli-

Table 1. Proportion Correct on Repeated Exam Items As a Function of Presentation Pattern

	2001 to 2003					2002 to 2003				
	2001 Class		2003 Class			2002 Class		2003 Class		
	M	*SD*	*M*	*SD*	n^a	*M*	*SD*	*M*	*SD*	n^a
Nonactive then active learning	$.63_a$	.26	$.79_b$	.16	9	$.64_a$	.22	$.79_b$	.11	10
Both active learning	$.92_a$	.02	$.93_a$	.02	4	$.85_a$	.12	$.82_a$	.13	19
Same nonactive learning	$.77_a$	.14	$.74_a$	.17	11	$.74_a$	.21	$.72_a$	.20	61
Different nonactive learning	$.77_a$	.18	$.77_a$	.11	14	$.75_a$	.19	$.74_a$	.18	15
Subtotal active learning	.92	.02	.84	.15	—	.85	.12	.81	.13	—
	$n = 4$		$n = 13$		—	$n = 19$		$n = 29$		—
Subtotal nonactive learning	.73	.18	.75	.14	—	.73	.20	.72	.20	—
	$n = 34$		$n = 25$		—	$n = 86$		$n = 76$		—
Total	.75	.18	.79	.15	38	.75	.19	.75	.18	105

Note. The subtotals combine active or nonactive items within year. Different subscripts across classes (i.e., across rows) indicate significant mean differences identified by post hoc *t* tests.
[a]Represents the number of repeated exam items.

cated across comparisons of two sets of classes, arguing for some generalizability of the results across different classes.

Using a much more limited set of exam items, our findings suggest that lost classroom coverage because of greater time devoted to active learning techniques need not negatively affect students' performance. Advocates for active learning (e.g., Bransford et al., 2000) might argue that the metacognitive framework, begun in class through active learning, might carry over to direct subsequent learning. This carry-over is exactly what Schwartz and Bransford (1998) intended to demonstrate when their active learning experimental condition followed students' active case analysis with lecture and reading. Their data did show deeper learning for these active learners whose understanding of subsequent materials was likely affected by their prior experiences with case analyses, but further research is needed to tease apart the impacts of lecture and readings as subsequent support materials. Our data, although limited, are consistent with the argument that materials not covered in class, but rather left to independent coverage outside class through assigned readings, need not be sacrificed and indeed may be compensated for by prior metacognition.

A liability of much quasi-experimental field research exploring the effectiveness of teaching techniques is the inability of researchers to randomly assign students to classes and include multiple instructors in each condition. Exam items and the materials they cover vary in their difficulty, and classes of students differ along many uncontrolled dimensions. Although the level of learning tapped by each exam item, as well as class GPA data, helped rule out some of these common alternative explanations, this study remains open to some of these threats to both its internal and external validity. Students self-selected into classes, and a single instructor was involved throughout (although she coded exam items retrospectively with the help of detailed class notes to minimize potential experimenter effects). However, patterns in the data, the replication of these patterns across two class pairings, and several companion comparisons of students' performance strengthen our conviction about the validity of our conclusion that active learning presentations enhanced students' performance on multiple-choice exams.

For instructors, the obvious implications of these findings are to target some class materials for active learning approaches and to employ supported readings in place of lecture for some topics. As in this study, decisions about when and how to present materials remain the prerogative of instructors and rest heavily on instructors' confidence in the quality and clarity of assigned readings. Deciding when to use any learning technique requires that instructors continually ask themselves what is most important in the materials they are covering and seek out information from students about which points remain muddled for them. (Item analyses of exams also may identify trouble areas, and indeed in this study, the instructor identified and targeted materials for active learning interventions that these data analyses subsequently identified as low scoring.) Because students' capability for remembering information is limited, encouraging instructors to identify, then target, the most central points in course materials is an essential step toward effectively exploiting all learning techniques.

Another key step is to expand one's arsenal of active learning techniques. The definition we used, "anything that involves students in doing things and thinking about the things they are doing" (Bonwell & Eison, 1991, p. 2), opens up a wide array of possibilities. These approaches can range from elaborate demonstrations, involving full class periods and structured activities, to simple extensions of techniques already used, such as lectures and videos. In addition to a rich array of specific creative and tested possibilities offered in *Teaching of Psychology*, there are some simple general parameters instructors might consider.

For example, following videos or lectures with minute papers (Angelo & Cross, 1993), where students jot down the most valuable thing they just learned and what remains muddled (then they can share these with their peer team or the class as a whole), can identify students' misunderstandings as well as coalesce their knowledge (Butler et al., 2001). Having students discover patterns in individual cases prior to a lecture prepares students to better predict the results of research studies (Schwartz & Bransford, 1998), suggesting that discussion and discovery prior to lecture or reading may prepare students to learn more. Framing discussions, both in small peer groups and among the class overall, using thought-provoking questions (see King, 1995) is another technique to encourage students' engagement. Completing a measurement scale can make a subsequent lecture more personally meaningful for students and encourages them to think critically and concretely about the measure and its underlying construct. Psychology classes offer especially rich and meaningful materials from which instructors readily can encourage more active learning in their classes, and our data argue that doing so will enhance classes' performance.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Pearson Education.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom* (ASHE–ERIC Higher Education Rep. No. 1). Washington, DC: The George Washington University, School of Education and Human Development.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Butler, A., Phillmann, K. B., & Smart, L. (2001). Active learning within a lecture: Assessing the impact of short, in-class writing exercises. *Teaching of Psychology, 28,* 257–259.

Halonen, J. S., Brown-Anderson, F., & McKeachie, W. J. (2002). Teaching thinking. In W. J. McKeachie (Ed.), *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (11th ed., pp. 284–290). Boston: Houghton-Mifflin.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). The effects of cooperative, competitive, and individualistic goal structure on achievement: A meta-analysis. *Psychological Bulletin, 89,* 47–62.

King, A. (1995). Inquiring minds really do want to know: Using questioning to teach critical thinking skills. *Teaching of Psychology, 22,* 13–17.

Lake, D. A. (2000). Active learning: Student performance and perceptions compared with lecture. In J. A. Chambers (Ed.), *Selected papers from the 11th International Conference on College Teaching and Learning* (pp. 119–124). Jacksonville: Florida Community College at Jacksonville.

Lonka, K., & Ahola, K. (1995). Activating instruction: How to foster study and thinking skills in higher education. *European Journal of Psychology of Education, 10,* 351–368.

Mathie, V. A., Beins, B., Benjamin, L. T., Jr., Ewing, M. M., Hall, C. C. I., Henderson, B., McAdam, D. W., & Smith, R. A. (1993). Promoting active learning in psychology courses. In T. V. McGovern (Ed.), *Handbook for enhancing undergraduate education in psychology* (pp. 183–214). Washington, DC: American Psychological Association.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59,* 14–19.

McKeachie, W. J. (2002). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (11th ed.). Boston: Houghton-Mifflin.

McKinley, N. M., & Hyde, J. S. (1996). The Objectified Body Consciousness Scale: Development and validation. *Psychology of Women Quarterly, 20,* 181–215.

Nelson, C. E. (1996). Student diversity requires different approaches to college teaching, even in math and science. *American Behavioral Scientist, 40,* 165–175.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–522.

Sivan, A., Leung, R. W., Woon, C., & Kember, D. (2000). An implementation of active learning and its effects on the quality of student learning. *Innovations in Education and Training International, 37,* 381–389.

Yoder, J. D. (1999). *Women and gender: Transforming psychology.* Upper Saddle River, NJ: Prentice Hall.

Yoder, J. D. (2003). *Women and gender: Transforming psychology* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Notes

1. We thank Tom Angelo for his helpful comments and learner-centered seminar on teaching, assessment, and learning.
2. Send correspondence to Janice D. Yoder, Department of Psychology, University of Akron, Akron, OH 44325–4301; e-mail: jyoder@uakron.edu.

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