

Examining exam reviews: A comparison of exam scores and attitudes

Jana Hackathorn¹, Kathryn Cornell², Amy M. Garczynski³, Erin D. Solomon³, Katheryn E. Blankmeyer³, and Rachel E. Tennial³

Abstract: Instructors commonly use exam reviews to help students prepare for exams and to increase student success. The current study compared the effects of traditional, trivia, and practice test- based exam reviews on actual exam scores, as well as students' attitudes toward each review. Findings suggested that students' exam scores were significantly higher for the exams following both a traditional review and a trivia review than for the exam following the practice test review. Immediately after the review, three attitude measures (i.e. confidence, feeling prepared, and helpfulness of review) were lowest after the traditional review. Finally, immediately after taking the exam, students reported higher ratings (i.e. confidence, feeling prepared, and helpfulness) after the trivia review. Implications of these results are discussed.

Keywords: Exam Reviews, Teaching Effectiveness, Learning

I. Introduction.

Using exam reviews to help students prepare for an exam leads to better test results in comparison to no review at all (King, 2010). However, a majority of the literature regarding exam reviews has been either anecdotal, or has mainly focused on measures of students' attitudes. For example, Middlecamp (2003) measured students' perceptions of a review game for an abnormal psychology course. The findings suggested that most of the students enjoyed the review game and felt it was challenging. However, Middlecamp did not measure the actual effectiveness of playing the game on exam scores. This is problematic, as student perceptions of the review's effectiveness may not align with the review's actual effectiveness. Furthermore, literature regarding direct comparisons of differing types of exam reviews is sparse. Arguably, any exam review, if done correctly, can garner positive effects for students' success. However, what remains unclear is whether certain exam reviews are more effective than others. Thus, the current study sought to directly compare the effectiveness of a few variations of exam reviews (i.e. traditional, trivia, and practice-test).

Traditional exam reviews, a common approach, ask students to compose questions that the instructor will answer during a scheduled, class-wide meeting. However, the success of a traditional exam review is contingent upon students coming to class prepared to ask questions. Moreover, the traditional exam review is often received by students as 'just another class lecture', passive and disengaging, and only including, or benefitting, the few students who prepared (King, 2010; Paul, Hollis, & Messina, 2006). In attempts to remedy this, many educators have created and experimented with new ways to implement reviews with the hopes of

¹ Jana Hackathorn, Murray State University, 209 Wells Hall, Murray KY 42071, jhackathorn@murraystate.edu

² Murray State University

³ Saint Louis University

improving exam scores, increasing attendance for review days, and also in influencing attitudes (e.g., satisfaction, feelings of efficacy) about the given subject.

One alternative is for instructors to implement trivia-style games, similar to the television show, *Jeopardy*, to review material. These are often utilized because students appear to enjoy them, and can be easily created for any level course regardless of material type or difficulty. For example, Keck (2000) created a trivia game, based on the game show *Jeopardy*, as a supplement to a traditional review before a final exam in an upper level chemistry course. Results indicated that students enjoy trivia-based review games, and that students who participated yielded higher exam scores than those who did not participate.

Similarly, Paul, Hollis, and Messina (2006) examined the effectiveness of a trivia game on exam scores. Their findings suggested that students' grades increased significantly across all four exams for students who played the trivia game, as opposed to students who chose not to play. Self-report measures indicated that the game experience taught students how and what to study. Perhaps most importantly, students reevaluated their study methods based on the inadequacies of their knowledge discovered during the game reviews.

A major limitation to Paul and colleagues' (2006) study, however, was that students chose whether or not to attend the class when the game was played. That is, students self-selected into the experimental or control group. This is problematic for a couple of reasons. First, it is unclear whether students who did not choose to play the game received a different type of review, or no review at all. Moreover, it is unclear whether this type of review is helpful for all students, or just the students who would choose to play a review game. That is, an individual difference variable, such as intrinsic motivation or interest in the subject material, could be producing higher exam grades. Perhaps those students who chose to play a trivia game would have scored higher on the exams regardless, because they possess greater motivation.

Finally, practice tests are an additional alternative to traditional reviews: however, past literature is inconclusive regarding their effectiveness. Some studies indicate that students prefer technology-based learning, particularly practice tests, because they appreciate the opportunity to check their knowledge level, and get a 'sneak peek' at the format and wording of exam questions (King, 2010; Pemberton, Borrego, & Cohen, 2006). For example, Balch (1998) assigned students to either take a practice test, or to review a completed practice test prior to an exam, and found that actively taking the practice test significantly improved subsequent exam scores. The author argued that the process of actually taking an exam, and subsequently getting answers incorrect, allows students to understand their weaknesses. Additionally, Balch argued that by taking the practice tests, the students are engaging in deeper cognitive elaboration, improving understanding and retention of material.

Additionally, King (2010) used a slightly different version of the practice test exam review, in which students used clickers to answer multiple choice questions. Attendance and exam grades were higher for clicker reviews relative to other reviews, however, similar to Paul and colleagues (2006), results are confounded by the fact that students self-selected into the exam review or no exam review conditions.

Conversely, other studies have shown little to no effects of the practice test review. Pemberton, Borrego, and Cohen (2006) found that using online computer-based practice tests in reviews did not improve exam grades compared to those who had traditional review. Additionally, Kalat (1983) showed that a practice test that mimicked the style of real exams had no effect on exam scores, as compared to a control group with no review.

In summary, past research is inconclusive regarding the effectiveness of different types of exam reviews. What does seem clear is that students report greater liking of non-traditional exam review types, but it is not clear whether this translates to higher exam scores. Thus, the current study compared the effectiveness of three types of exam reviews. Although there are a plethora of exam reviews available, the traditional, trivia, and practice test exam reviews were chosen as a focus for the current research because they are commonly used in college classrooms and there is inconclusive evidence as to their effectiveness. It is important to note that the purpose of the current study was not to discredit any type of exam review. Instead, as there are limitations in the past literature as well as contradictions in success rates and effectiveness, the current study sought to compare the effects of a traditional review, trivia game, and practice-tests on actual exam scores. In addition, the current study sought to measure students' attitudes toward each review, and used a within-groups design, utilizing only students who attended all three types of review as a means of comparison.

As this study is one of the first known to purposefully compare various types of review, we posed the following research question: which type of exam review will result in the highest exam scores? As little past literature has compared the effectiveness of these types of exam reviews, it was difficult to predict which would result in the highest exam scores. Therefore, no specific hypotheses were made as to which exam review type would be the most effective for actual exam scores. Additionally, we were interested in which type of review style would be rated most favorably by students. Based on prior literature (e.g., King, 2010), we expected that students' overall attitudes would be least favorable of the traditional exam review immediately following the review, and again after the actual exam was taken. Specifically, we expected students to rate the traditional exam review as the least helpful, report feeling the least confident and the least prepared relative to the practice-test or trivia review styles.

II. Methods.

A. Participants.

Participants in the current study were 78 undergraduates enrolled in two separate sections of a social psychology course, each with a different instructor, at a Midwestern university. Both courses met three days a week, for 50 minutes each, during a traditional 16-week semester. However, as it is consistent with the purpose of the current study, only those students who participated in all three of the exam reviews were included in the analysis of actual exam scores ($n = 44$). Those students' ages ranged from 18-22 years ($M = 19.75$, $SD = 1.54$, $Med = 20.00$). The classes contained Freshmen ($n = 7$), Sophomores ($n = 17$), Juniors ($n = 7$), and Seniors ($n = 5$). Eight students did not provide their college year. Additionally, there were 26 women, 10 men, and eight participants who chose not to answer. The ethnic background of students was predominantly Caucasian ($n = 27$), but also included African-American ($n = 3$), Indian ($n = 2$), Hispanic ($n = 1$), Bi-racial ($n = 1$), Asian ($n = 1$), and Bosnian ($n = 1$) ethnicities. Eight students' ethnic information was not reported. Finally, most of the students were psychology majors ($n = 14$), however there was also language majors ($n = 5$), biology ($n = 2$) and human resources ($n = 1$). The remaining 23 students did not answer the question.

B. Procedure.

Undergraduate psychology students in two upper level social psychology courses were instructed at the beginning of the semester that throughout the duration of the course an examination of various teaching techniques and exam reviews would occur. Students were also informed that any and all activities were voluntary, and they could choose not to complete any or all of the surveys. They did not receive any extra credit for their participation or completion of surveys. However, in an attempt to increase attendance in each of the exam reviews, extra credit was offered for participating in each exam review.

To control for variance across courses, the instructors worked together to cover the same material using the same notes and identical exam review procedures. Additionally, both courses used the same textbook, the same homework assignments, and assessed learning through the exact same multiple choice exams. On the class period prior to each of the three exams, students were presented with a different type of exam review. The reviews consisted of a trivia-style game review for the first exam, followed by a traditional review for the second exam, and finally a practice test prior to the third exam, which was also the final exam. To increase attendance and participation at each of the exam reviews, extra credit points were added to students' final grades. In order to maintain the integrity of the dependent variable, exam scores before the extra credits were added were used in the analyses.

1. Trivia Exam Review.

On the scheduled class day prior to the first exam, students participated in a trivia game in which they were randomly assigned to teams of five to six students. The trivia game was set up so that the instructors created a series of questions prior to the class. Questions ranged from giving definitions and having students recall or recognize (via a multiple-choice format) the correct concepts, to giving real-world examples and students correctly applying concepts. Students took turns representing their group for each question. Groups were asked questions pertinent to material on the upcoming exam, and a point was awarded to the first team with the correct answer.

2. Traditional Exam Review.

On the scheduled class day prior to the second exam, students participated in a traditional review. Students were instructed to come to class with questions regarding material that was unclear or needed further clarification. During the class period, students were called on to read a prepared question aloud and the instructor would answer the question for them, or provide clarification regarding subject matter that was unclear. Other students were not necessarily given an opportunity to answer the questions, but were allowed to ask follow-up questions or discuss the material until everyone in the class understood. There were no limits to how many questions each student could ask, and instructors used the full class time.

3. Practice Test Exam Review.

On the scheduled class day prior to the third exam, students participated in a computerized practice test on their own computers outside of the classroom. The computer practice test took

place during the same 50 minute time period that the scheduled class occurred. Students were informed that the amount of questions they answered correctly would not impact their course grade or extra credit. The 20 questions on the practice test mimicked the format, wording, and material of actual multiple choice questions of the upcoming exam, but were not identical to actual test items. Only 20 questions were given, as this provided students the opportunity to take notes, look up answers if necessary, or get further clarification from their notes or textbooks. After students finished the practice test, they were informed of which question(s) they got incorrect, the correct answer(s) and an explanation of the correct answer(s).

C. Measures.

1. Exam scores.

Each test consisted of a variety of multiple choice questions, short answer questions, and essay questions. However, only the multiple choice portion of each of the three exams was used to measure effectiveness of exam reviews. This was done to reduce instructor bias in the measurement of the dependent variable. Analyzing only the multiple choice portion reduces teacher bias in grading, providing a more objective measure. Students' answers were either correct or incorrect. No partial credit was awarded. The first exam contained 16 multiple choice questions, the second exam contained 23 multiple choice questions, and the third exam contained 21 multiple choice questions. Since each exam had a different number of multiple choice items, total scores were converted into percentages. Each exam was identical across sections of the course, and care was taken to make the three exams as comparable as possible. Each exam contained questions at the Knowledge, Comprehension, and Application levels of Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956).

2. Attitudes Following Exam Reviews.

Immediately following each exam review, students were given three-single-item surveys assessing their exam confidence, feelings of preparedness, and the helpfulness of each exam review. Participants responded to three questions (i.e., "After this exam review, how confident do you feel about taking the exam?" "After this exam review, how prepared do you feel?" and "How helpful do you feel this exam review was?" on a 7-point Likert scale ranging from (1 = *not at all* to 7 = *very much*).

A slightly revised version of the attitude survey was administered again immediately after students completed the exam. This provided students an opportunity to reevaluate their perceptions of each review style, after having seen the items on the exam. Specifically, the three items were reworded to indicate tense (e.g., after taking this exam, how helpful do you feel this exam review was?").

III. Results.

As there were two instructors in the course, analyses were conducted to determine if there were any differences in mean scores across instructors. Results indicated that there were no significant differences in scores for any of the three exams. Thus, for the sake of brevity, the data was collapsed across the two classrooms.

To examine our main research question, which queried the effectiveness of each type of review on actual exam scores, a repeated measures ANOVA was conducted using only exam scores for the students who participated in all three exam reviews ($n = 44$). Results indicated there was an effect of review, $F(2, 86) = 30.93, p < .001, \eta^2_p = .43$. Pairwise comparisons indicated that exam scores after traditional review ($M = .87, SD = .09$) and trivia reviews ($M = .86, SD = .10$) were significantly higher than after the practice-test-based review ($M = .74, SD = .10, ps < .001$). Traditional and trivia exam scores were not significantly different from one another ($p > .05$). See Figure 1 for an illustration of the differences between exam scores.

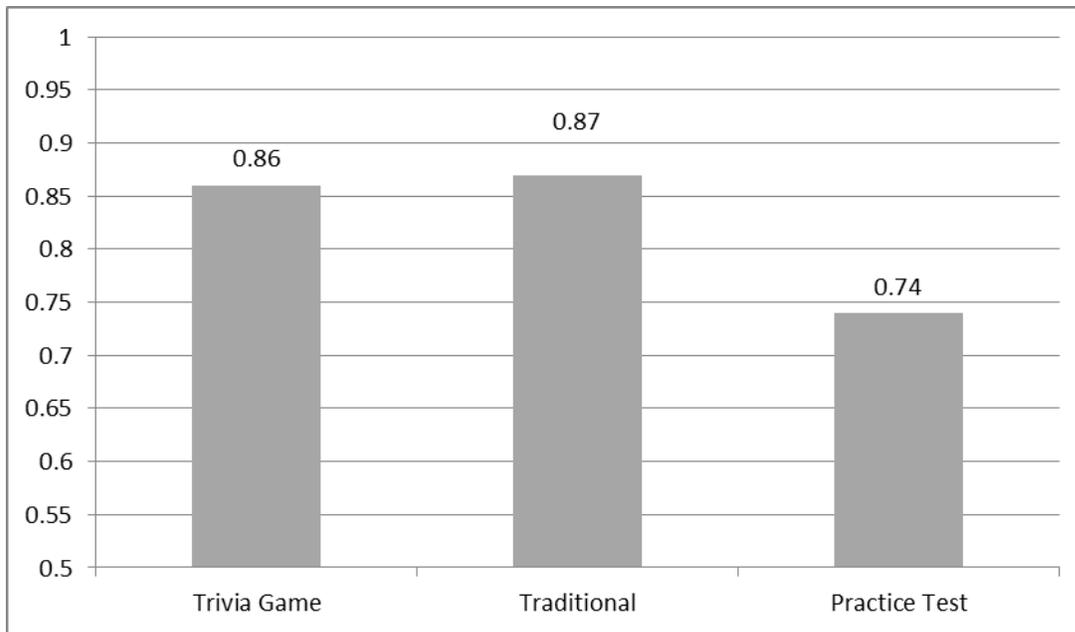


Figure 1. Percentage of correct answers between three exam review types.

In addition, it was hypothesized that attitudes toward the traditional review would be less favorable than the trivia or practice test reviews. Repeated measures ANOVAs were conducted on students' ratings of confidence, preparedness, and how helpful they felt the exam reviews were, immediately after the exam review. Findings indicated that, although the attitudes following the trivia review and the practice-test review were not significantly different from one another ($p > .05$), students reported more confidence after the trivia and practice-test review than after the traditional review ($p = .002$ and $p < .001$, respectively). Students also reported feeling more prepared after the practice-test review than both the trivia ($p = .019$) and the traditional ($p < .001$) review. Finally, students reported that both the trivia and practice-test review were more helpful than the traditional review ($p < .001$); however, student ratings of helpfulness for the trivia and practice-test reviews were not significantly different from one another ($p > .05$). See Table 1 for summary of analyses for each item.

Immediately following each exam, students rated their confidence again regarding each of the reviews. Results of repeated measure ANOVAs indicated students reported more post-exam confidence following the trivia review than after the practice-test review ($p = .003$), but ratings were not different from the traditional review ($p > .05$). Additionally, post-exam confidence was marginally higher after the traditional review than the practice-test review ($p = .075$). Students also reported feeling more prepared after the trivia and traditional review than

after the practice-test review ($p = .014$ and $p = .005$, respectively). However, the trivia and traditional review were not significantly different from one another in students' feelings of preparedness before the exam. Finally, after taking the exam, students reported the trivia game was more helpful than the traditional review ($p = .009$) but was not significantly different from the practice-test review ($p > .05$). Additionally, the traditional review and the practice-test review were not significantly different from one another ($p < .05$) in the students' reports of helpfulness. See Table 2 for summary of analyses for each individual attitude item.

Table 1. Analysis of Attitude Items Immediately after the Exam Review.

	Trivia M(SD)	Traditional M(SD)	Practice-test M(SD)	<i>F</i>	<i>p</i>	η^2_p
Confidence	4.64(1.11) ^b	3.98(1.03) ^{ac}	4.74(0.87) ^b	9.82	<.001	.18
Prepared	4.11(1.09) ^c	3.72(1.17) ^c	4.60 (0.97) ^{ab}	8.86	<.001	.16
Helpful	5.39(1.11) ^b	3.89(1.40) ^{ac}	5.28(1.26) ^b	20.74	<.001	.32

Note: Superscript notations indicate significant difference from other review; ^a = difference from trivia, ^b = difference from traditional, ^c = difference from practice-test.

Table 2. Analysis of Attitude Items after the Exam.

	Trivia M(SD)	Traditional M(SD)	Practice-test M(SD)	<i>F</i>	<i>P</i>	η^2_p
Confidence	5.56(1.19) ^c	5.24(1.08)	4.79(0.98) ^a	5.28	.007	.14
Prepared	5.38(1.16) ^c	5.41(1.17) ^c	4.65(1.36) ^{ab}	5.52	.006	.13
Helpful	5.22(1.38) ^b	4.54(1.22) ^a	4.70(1.55)	2.97	.057	.08

Note: Superscript notations indicate significant difference from other review; ^a = difference from trivia, ^b = difference from traditional, ^c = difference from practice-test.

IV. Discussion.

The current study is the first known to directly compare the effectiveness of traditional, trivia, and practice-test reviews on actual exam scores. Findings indicate that exam scores were significantly higher following both a traditional and a trivia review than following the practice-test review. This suggests that compared to practice-test reviews, traditional and trivia exam reviews more effectively increase students' exam scores. These findings support and extend past literature which indicates that not only do students enjoy trivia games as an exam review, but that students who actively participate in trivia games tend to show increased exam scores (e.g., Keck, 2000; Paul, Hollis, & Messina, 2006), especially when compared to no review at all. Our findings also support the notion that practice tests have very little effect on exam scores (e.g., Kalat, 1982), although, it is important to note that practice tests reviews are better than no review at all (Balch, 1998).

Additionally, the current study examined students' attitudes toward each of the exam reviews. Based on findings in prior literature in which students' attitudes were measured, we expected that students' attitudes would be most favorable of the trivia and practice test reviews, and least favorable of the traditional review (King, 2010; Paul et al. 2006). That is, we expected students to rate the traditional exam review as least helpful, and would report they felt less confident, and less prepared than after the practice-test or trivia reviews. Our findings partially

supported this hypothesis. Results indicated that prior to taking the exam, all three attitude measures (confidence, feeling prepared, and helpfulness of review) were lowest after the traditional review, in spite of the fact that the traditional review was more helpful than the practice test. However, after taking the exam, the students' attitudes shifted in that they were least favorable of the practice-test-based review. Perhaps students are aware of which types of reviews are actually the most helpful to them, but only after taking the actual exam and not immediately after the review itself. Future studies should investigate if students are aware of their learning and studying styles, and whether this corresponds with the types of reviews they prefer.

At both points, immediately after the review and after taking the exam, students' ratings of the trivia game were the most favorable of any of the review styles. Importantly, students' scores on the actual exam were also highest after the trivia review. There are several possible reasons for this finding. Perhaps the trivia game reviews are more appreciated by students because they are similar to practice test reviews which give students a 'sneak-peek' at test wording while also explicitly pointing out areas where more studying is needed (King, 2010). Or, perhaps as students find the trivia games more fun they feel more at ease and less anxious (Middlecamp, 2003). This lack of anxiety may help them to study more effectively and increase their chances of success. However, measuring students' fun and anxiety toward the review was beyond the scope of this study.

Our findings also suggest that a small disconnect exists between the perceived effectiveness and the actual effectiveness of the traditional exam review. That is, students did not seem to realize the helpful nature of traditional exam reviews. Immediately following a review, students feel that the traditional reviews are the least helpful, but their actual scores were higher after the traditional review than the practice-test review. Moreover, even after taking the actual exam, students still did not perceive the effectiveness of the traditional review. It has been noted that the success of the traditional exam reviews are contingent upon students attending the session prepared with questions for the instructor, which suggests some amount of studying before the review session (King, 2010; Paul, Hollis, & Messina, 2006). It is possible that students perceive that the traditional review session is supposed to guide their studying or even replace their studying altogether. Thus, a lack of preparedness and student attitudes may lead to the perceived ineffectiveness of traditional reviews. Another possibility is that the students who feel the least prepared following the review session put additional effort into studying for the exam, thereby explaining the disconnect between feelings of preparedness and actual exam scores.

Past studies (e.g., King, 2010) have suggested that perhaps students find traditional exam reviews to be stuffy, awkward, or just plain boring. As part of this study, although not part of our explicit analysis, students were asked open-ended questions regarding what they liked and disliked about each exam review. After the traditional exam review, a theme emerged within students' comments pertaining to the dislikes of the traditional review stating that the traditional review was boring, an overall feeling that not all of the material was covered, and that the overall session was disorganized. As students generate their own questions to a traditional exam review, and ultimately take turns, the material mentioned in one question is not necessarily connected to the question that was asked before it. This may inhibit proper note taking, and even lead the material to become more confusing to them. Thus, they feel less prepared and confident, and ultimately underestimate the helpfulness of the review session. Perhaps if instructors structured the traditional exam reviews so that students' questions were grouped by chapter or concept, this might ease students' concerns.

V. Limitations.

Although this study adds to the literature regarding exam reviews, it is not without limitation. As would be important in any class, care was taken to ensure the exams were of similar difficulty. However, as each exam covers different topic material (e.g., self vs. influence), it is possible that there were inherent differences in difficulty. Thus, it is possible that the exam following the practice-test review was more difficult than the exam following the traditional and trivia reviews. A brief qualitative examination indicated that all three exams contained multiple choice questions at the knowledge, comprehension, and application levels of Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). However, it should be noted that the exam following the practice-test review contained four application level questions, while the traditional review exam contained only two application questions and the trivia exam contained only one. Although there is a slight discrepancy, it is our belief that this is not a noteworthy enough difference to sway scores in such an extreme direction.

VI. Recommendations for Future Research.

This study offers some important information for the scholarship of teaching and learning regarding the actual outcomes of exam review sessions, as well as some of students' perceptions of the effectiveness of each exam review style. However, it does not provide much qualitative information regarding what students may or may not appreciate about each type of review. Knowing the unique features that students like and dislike in an exam review, might help instructors to develop a review session that is not only effective, but also highly appreciated and enjoyed by students.

As instructors, we should use these findings, in addition to future findings, to create review sessions that students can simultaneously benefit from and enjoy. Perhaps, combining the different types of exam reviews to create one that is both effective and enjoyable is the most beneficial type of review. For example, each review could begin with a trivia review game that takes only half the allotted class time, and then cover the most 'missed' topics in a more traditional review fashion. Or, after each trivia question is answered, probing students for information regarding what part of that specific material is still muddy, or providing more information as to why the answer is the correct answer. This conglomeration of review styles could significantly ease students' anxiety, increase their enjoyment and ultimately their participation in exam reviews, and importantly increase their knowledge of the material which culminates in increased exam scores.

References

Balch, W.R. (1998). Practice versus review exams and final exam performance. *Teaching of Psychology, 25*, 181-184.

Bloom, B.S., Engelhart, M.D., Furst, F.J., Hill, W.H., and Krathwohl, D.R. (1956). *Taxonomy of educational objectives: Cognitive domain*. New York: McKay.

Drouin, M.A. (2010). Group-based formative summative assessment related to improved student performance and satisfaction. *Teaching of Psychology, 37*, 114-118.

Kalat, J.W. (1983). Effect of early practice test on later performance in introductory psychology. *Teaching of Psychology, 10*, 53.

Keck, M.V. (2000). A final exam review activity based on the Jeopardy format. *Journal of Chemical Education, 4*, 483.

King, D. (2010). Redesigning the preexam review session. *Journal of College Science Teaching, 40*, 88-96.

Middlecamp, M.K. (2003). Uncover the disorder: A review activity for abnormal psychology courses. *Teaching of Psychology, 30*, 62-63.

Paul, S.T., Hollis A.M., and Messina, J.A. (2006). A technology classroom review tool for general psychology. *Teaching of Psychology, 33*, 276-279.

Pemberton, J.R., Borrego, J., and Cohen, L.M. (2006). Using interactive computer technology to enhance learning. *Teaching of Psychology, 33*, 145-147.