Journal of Behavioral and Neuroscience Research
Volume 9, Issue 2, Pages 151-157

Effects of Test-Enhanced Learning in a Cognitive Psychology Course

Shanta Hattikudur & Bradley R. Postle
University of Wisconsin – Madison

Given recent findings that test-enhanced learning (TEL) improves students' understanding and retention of material through regular quizzing (e.g., Lyle & Crawford, 2011; McDaniel, Roediger, & McDermott, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011), we incorporated TEL-inspired online quizzing into a cognitive psychology undergraduate lecture course. With the addition of quizzing to the course, class performance improved according to several criteria relative to semesters when quizzing was not used. Quizzing was found to be a practical way to help students learn and retain the course material, without taking class time.

Keywords: *learning*, *quizzing*, *testing*, *undergraduate*

A primary goal for any academic course is to help students learn and retain information. Instructors advocate and students engage in various studying tactics in order to best achieve this goal, including rereading the text, recopying notes, and reviewing slides from lecture. Recent research in cognitive psychology has revitalized interest in another approach: Quizzing can promote long-term retention of course material even more effectively than does spending a comparable amount of time studying the material in more conventional ways (Lyle & Crawford, 2011; McDaniel, Roediger, & McDermott, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011; Roediger & Karpicke, 2006). In cognitive psychology, this phenomenon is variously referred to as "test-enhanced learning" (TEL), or the "testing effect". In one example, Butler and Roediger (2007) found that long-term memory for lecture material increased by 31% for students who took short-answer quizzes on this material relative to students who studied a summary of the material, and by 135% relative to students who did not review the material at all. Because quizzing provides greater benefits to retention than does restudying the material, it is believed that the mental process of actively retrieving information from memory is the mechanism driving the testing effect (e.g., Carpenter & DeLosh, 2006; Carrier & Pashler, 1992; McDaniel & Masson, 1985; McDaniel, Roediger, & McDermott, 2007). Further evidence for this idea comes from the fact that TEL is strongest, and longer lasting, for short-answer quizzes relative to multiple-choice quizzes. (The former require free recall and can elicit a deeper level of processing and/or elaboration, whereas the latter can often be answered based on familiarity.)

Immediate feedback during quizzing also improves learning, especially for students with lower initial knowledge (Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007). Additionally, it is important to provide corrective feedback to prevent individuals from retaining long-term memories of any incorrect information that they may produce (or, with multiple choice, select) during quizzing (e.g., Kulhavy, 1977; Pashler, Cepeda, Wixted, & Rohrer, 2005). That is, producing/selecting incorrect information during quizzing can create a false memory. In addition to its beneficial effects in learning and retention, frequent quizzing may also help students to monitor their understanding of the material and encourage them to study course material on a regular basis. Thus, a growing body of research in cognitive psychology indicates that quizzing, although traditionally used for assessment purposes, can also be an effective tool for improving student retention of course material.

Prompted by the findings reviewed here, the authors incorporated TEL-inspired quizzing into a large-lecture course on Cognitive Psychology taught at the University of Wisconsin-Madison. The Spring semester of 2010 was the first time that TEL-based quizzing was implemented in this course, and this report gives a brief overview of its implementation and its effects on student performance in the course.

Journal of Behavioral and Neuroscience Research
Volume 9, Issue 2, Pages 151-157

Method

The TEL component of the course was a quiz that was assigned as homework after each of the course's 24 lectures. Quizzes were administered via the course Web site, which is supported by the university ("Learn@UW"). The format and procedure for quizzing took into account two principles of TEL: 1) short-answer quizzing is more effective than multiple choice quizzing (greater improvement and longer-lasting learning; McDaniel, Roedinger, and McDermott, 2007); and 2) unless feedback is given near-instantaneously after a question is answered *incorrectly*, there is a risk that the incorrectly produced (or selected) information is what will be retained.

The course met twice weekly for 75-minute lectures with no recitation sections. After each lecture, a quiz was posted online for students to complete outside of class. Each quiz comprised 2-3 "subquizzes", each relating to a main point of the associated lecture. Each subquiz began with one short-answer question followed by 2-3 related multiple-choice questions. The rationale for this arrangement was that whereas we wanted students to receive the superior benefit of having to answer short-answer questions, there was no practical way, via the Web-based interface, to evaluate, in real time, the accuracy of the content of the short-answer responses and, when necessary, to provide real-time corrective feedback. The associated multiple-choice questions, which could be scored automatically and in real-time by the Web site's software engine, were therefore intended to catch incorrect knowledge about that subquiz's concept. (Although students knew that short-answer questions were not graded, they were told that the course teaching assistant would periodically "spot check" a random sampling of quizzes, to ensure compliance with the quizzing procedure. Incidentally, the short answer questions were also recycled as the basis for pre-exam review sessions.)

Upon completion of a subquiz, incorrectly answered multiple-choice questions were flagged, and corrective feedback (ranging from one sentence to a short paragraph of text) was displayed for each incorrect answer (i.e., corrective feedback was prepared for each incorrect multiple-choice answer prior to the quiz being made available to students) (see Appendices A and B for screen shots of example questions and feedback). Once the student clicked a button in the corrective-feedback screen, s/he was returned to the beginning of that subquiz, and had to retake it in its entirety. This process repeated until every multiple-choice question in a subquiz was answered correctly, at which time the student could advance to the next subquiz. A student could pause partway through a quiz and return to it later as often as s/he liked (e.g., logging out and returning to it on a subsequent day was permissible); however, the student did not receive credit for having completed the quiz until each subquiz was completed with no errors.

Because the purpose of instantiating TEL-inspired quizzing was to help students better learn the material presented in class, the proximal outcome measure of this intervention was the mean and variance of student grades, relative to class performance in previous years when quizzing was not implemented. Although this was necessarily a post hoc comparison, its interpretability was enhanced by the fact that the comparison classes were taught by the same lecturer (B.R.P.), with the same core set of lecture slides and the same core set of exam questions. (See Appendix C for more detail about year-to-year continuity of content and exams.) The comparison classes were taught in the Spring semesters of 2001, 2002, 2003, 2004, 2006, and 2008. Although the final grades assigned to students were letter grades (the A–F scale that is standard at U.S. universities), cross-year comparisons were performed on numeric grades (total points possible = 100). Final grades in the course were calculated from 3 equally weighted, noncumulative exams that each covered one third of the material presented in class (75%), and completion of outside-of-class assignments (25%; in 2010 this corresponded to timely completion of assigned quizzes, in all previous years it corresponded to timely completion of assigned Web-based demonstrations of cognitive psychology experiments). Each year, the mapping from numeric to letter grades was fixed on the following scale: A ≥ 93%; AB ≥ 90%; B ≥ 80%; BC ≥ 75%; C ≥ 65%; D ≥ 50%.

Results

In the Spring 2010 semester, the Cognitive Psychology course had a final enrollment of 171 students. For the previous 6 times it was offered it had a mean enrollment of 137.3 (*range* = 114-183). As illustrated in Figure 1, the average numerical final grade for the six years that the course had been taught prior to the introduction of TEL-based quizzing was 80.3 (*range* = 79.1-82.0; average *SD* = 9.4). For Spring 2010, the mean performance was 84.0 (*SD* = 7.7). This improvement in the overall mean for 2010 vs. the preceding six years was highly significant: $t(6) = 28.8$; $p < .000005$.

Journal of Behavioral and Neuroscience Research
Volume 9, Issue 2, Pages 151-157

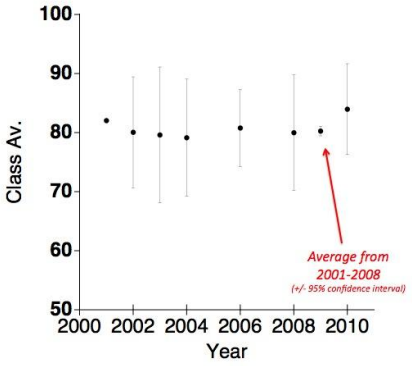


Figure 1. Class averages (mean +/- SD).

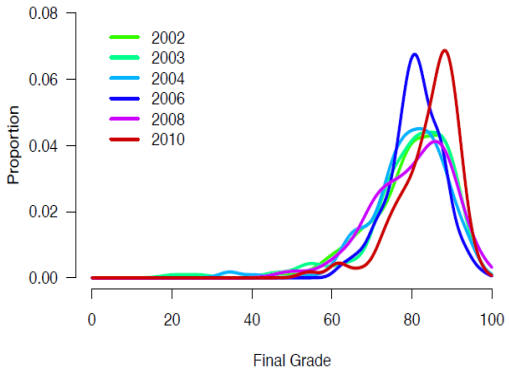


Figure 2. Distribution of final grades, by year. Histograms of each year's distribution of final numerical grades, transformed into density plots using the density.default feature in R. The algorithm disperses the mass of the empirical distribution function over a regular grid of 512 points and then uses the fast Fourier transform to convolve this approximation with a discretized version of the kernel and then uses linear approximation to evaluate the density at the specified points (Sheather & Jones, 1991; Silverman, 1986). Note that 2010 and 2006 were the two years with the highest enrollment.

Perhaps more diagnostic of the extent to which the TEL intervention improved learning in this course is the effect that it had in tightening up the distribution of scores across the class, particularly in shrinking the skew of lower-scoring students that is evident both from the distributions of final numerical grades, illustrated in Figure 2, and in the representative distributions of class performance from 3 years on Exam 1, illustrated in Figure 3. For the six offerings of the course prior to 2010, an average of 8.7% received a grade of D or F for the course; in 2010, 4.9% received a grade of D or F, $\chi^2(1, N = 879) = 7.943$, p = .005.

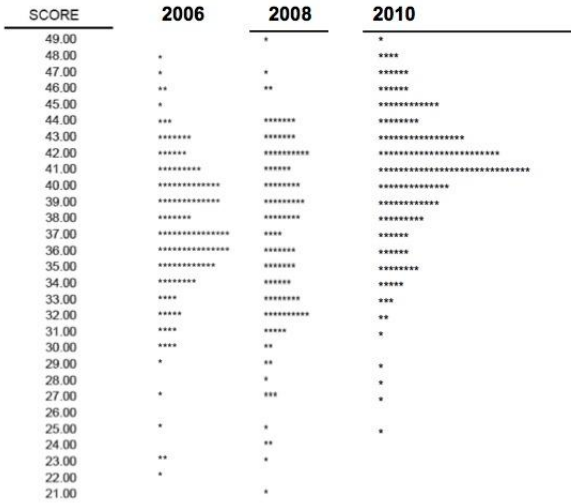


Figure 3. Histograms of grades on Exam 1 from three years (50 questions, total possible score = 50).

Finally, anecdotally, it is the professor's impression that the regular administration of quizzes fostered much more outside-of-class substantive dialogue among students and between the students and the TA and professor than has been typical of previous instantiations of the course.

Discussion

Our assessment is that the TEL-based quizzing intervention appeared to be highly successful. In terms of grades, class performance improved according to several criteria relative to semesters when quizzing was not used. And although the link is less direct, it is also possible that the high final enrollment (i.e., low attrition) for the class in 2010 reflects positive endorsement by students of this aspect of the class. That is, although initial enrollment in the course is typically between 180-200 students each year, a sizeable number drop the course during the 4-week add/drop period at beginning of the semester. Thus, final enrollment can be taken as an indirect index of the effectiveness of the class as judged by student behavior. Note that this speculation also calls attention to a possible placebo effect, in that from the first day of the class, when the procedure for quizzing was introduced, the rationale for introducing TEL-based quizzing was explained to them, and throughout the semester students often received the message from the lecturer and TA that our sincere intent was that they learn the material as well as possible. Put another way, it may be the case that because the quizzes were not used as a means of assessment, students recognized that the quizzing

feature of the course was in their best interest. The explicit knowledge that the TEL intervention was intended to improve academic performance may, in and of itself, have influenced performance. It is also worth mentioning that although the rationale for introducing TEL-based quizzing is supported by theories in cognitive psychology (and therefore relevant to the students in a cognitive psychology course), with proper explanation, we believe that students will recognize the merits of quizzing in any course, regardless of the content. We also believe a schedule of regular quizzes helped students by enforcing a "spaced" (vs. "massed) study schedule.

Overall, we found that the addition of a TEL-based quizzing feature to an undergraduate course can be useful in improving students' retention of the material as well as helping students to monitor their understanding of the material. From the practical standpoint of administering this scheme, Web-based quizzing has several appealing features. First, it does not take up valuable class time, because the quizzes are taken by students outside of class time (and so is conducive to any type of course – lecture, seminar, online, etc). Second, it does not take up valuable outside-of-class grading time, because grading is automated. Therefore, quizzing is relatively easy to add to a course and is likely to result in numerous benefits to student learning.

Author Note

First author's current affiliation: Temple University. Correspondence should be addressed to the first author at shanta.hattikudur@temple.edu. The authors thank Adam Riggall and David Sutterer for help preparing Figure 2.

References

Butler, A. C. & Roediger, H. L., III. (2007). Testing improves long term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4), 514-527.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition, 34*(2), 268-276.

Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition, 20*(6), 633-642.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4), 528-558.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*, 211-232.

Lyle, K. B. & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology, 38*(2), 94-97.

McDaniel, M. A. & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(2), 371-385.

McDaniel, M. A., Roediger, H. L., III, McDermott, K. B. (2007). Generalizing test-enhanced learning from laboratory to the classroom. *Psychonomic Bulletin and Review, 14*(2), 200-206.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3-8.

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied.* Advance online publication. Doi: 10.1037/a0026252

Roediger, H. L., III & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.

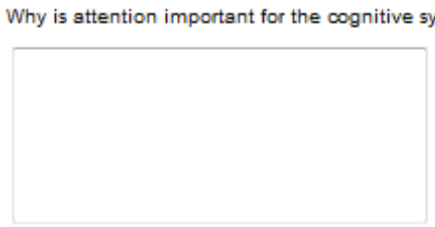
Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society* B, 683-690.

Silverman, B. W. (1986). *Density Estimation.* London: Chapman and Hall.

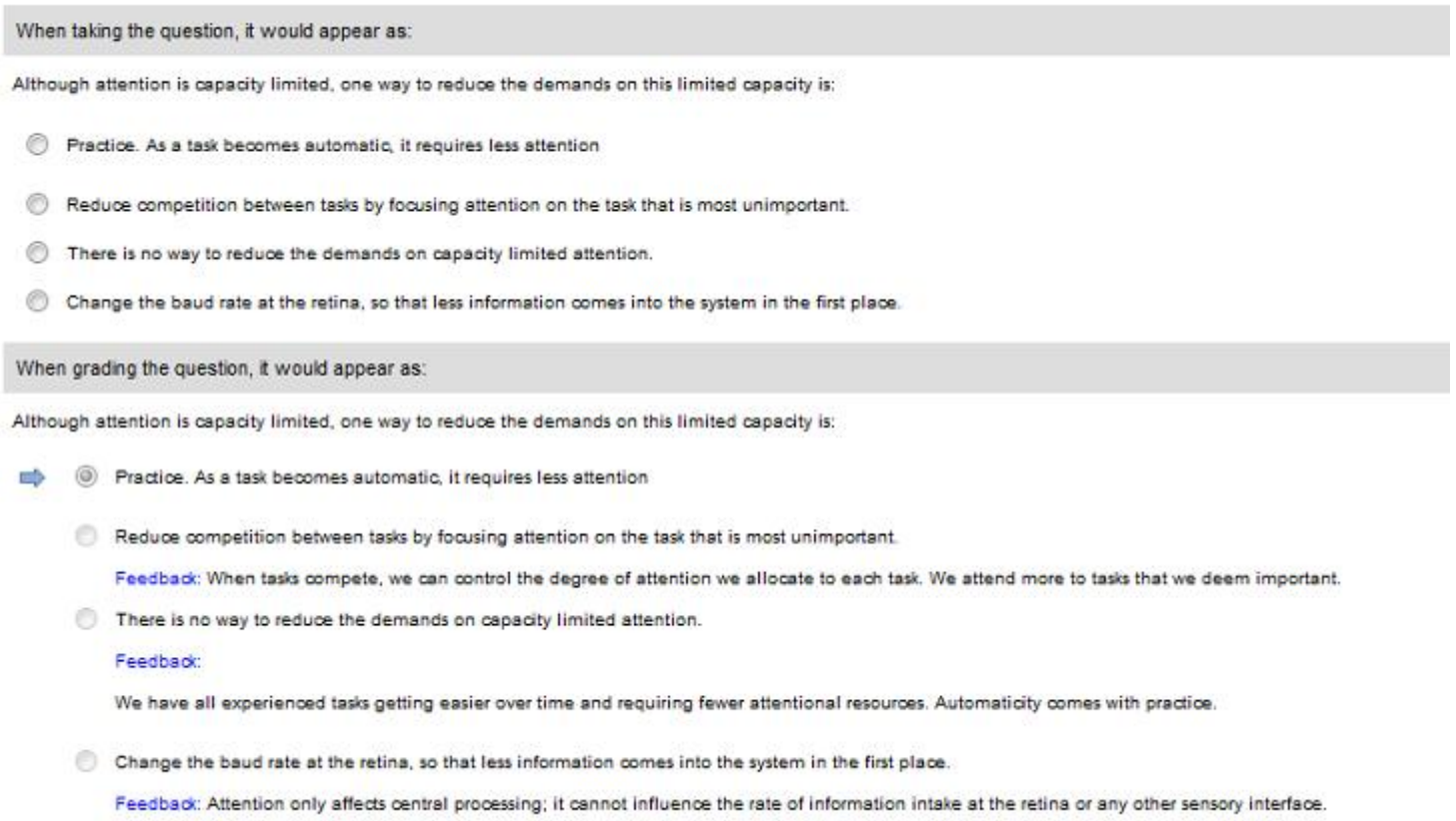
Journal of Behavioral and Neuroscience Research
Volume 9, Issue 2, Pages 151-157

Appendix A

Screenshot of short answer question.



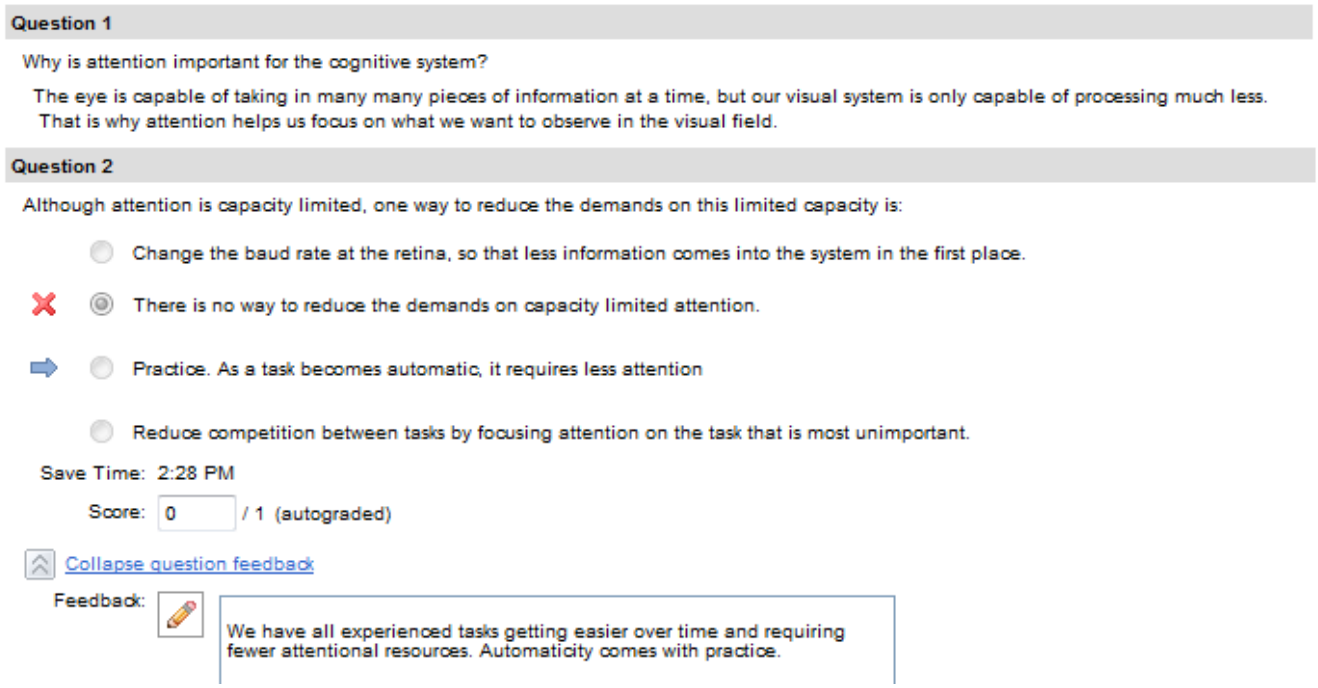
Screenshot of multiple-choice question with feedback.



Journal of Behavioral and Neuroscience Research
Volume 9, Issue 2, Pages 151-157

Appendix B

Screenshot of short answer question and multiple-choice question with feedback.



Journal of Behavioral and Neuroscience Research
Volume 9, Issue 2, Pages 151-157

Appendix C

Year-to-year continuity. Each time the course was offered the lecturer "tweaked" his lecture slides from the previous time he had taught the course. His informal recollection is that the changes were primarily stylistic, rather than substantive. Each year, however, the content of one of the course's 24 lectures was different, because this lecture was given by the course TA, and thus reflected her/his research interests and/or a topic with which s/he felt most comfortable. In 2010, this lecture was entitled "Gesture"; in 2008, "Language Acquisition"; in 2006, "Developmental Disorders" in 2004 ,"Auditory and Speech Perception; Structure Of Language"; in 2003, "Language Acquisition"; in 2002, "Auditory and Speech Perception; Structure Of Language"; and in 2001, "Language and the Brain". It was also true of exam questions that these were by-and-large the same each year, with the wording of a question sometimes changed for subsequent years if performance in a given year indicated that the question was answered either correctly or incorrectly by a disproportionate number of students. A small proportion of questions was also changed each year to reflect variation in the material that was presented that year. The largest such source of heterogeneity was the textbook: Across the seven years that the class has been taught to date, it has used four different text books.