Psychonomic Bulletin & Review
2007, 14 (2), 200-206

Generalizing test-enhanced learning from the laboratory to the classroom

Mark A. McDaniel, Henry L. Roediger III, and Kathleen B. McDermott
Washington University, St. Louis, Missouri

Test-enhanced learning refers to the fact that taking an initial test on studied material enhances its later retention relative to simply studying the material and then taking a final test. Most research on the testing effect has been done with materials such as word lists, and the general finding has been that the benefits of testing are greater when the initial test is a recall (production) test rather than a recognition test. We briefly summarize three experiments that extend these results to educationally relevant materials, namely brief articles, lectures, and materials in a college course. All three experiments demonstrated a robust testing effect and also revealed that an initial short-answer test produced greater gains on a final test than did an initial multiple-choice test. Furthermore, one experiment revealed a positive effect of immediate feedback given with the initial test. The educational implications are that production tests (short answer or essay) and feedback soon after learning increase learning and retention. In addition, frequent testing probably has the indirect positive effects of keeping students motivated and leading them to space out periods of study.

Educators generally view testing as a procedure to assess students' learning and to assign grades; consequently, educational researchers have focused on the analysis and development of testing as an evaluative technique, and thus have been concerned with issues such as test reliability and validity. In contrast, memory researchers have focused on the direct mnemonic effects of testing: Does testing material alter the knowledge that was tested? Early in the 20th century, researchers showed that testing (often called *recitation* in the early literature) improved later retention (e.g., Gates, 1917; Jones, 1923/1924). The improved performance on a later retention test arising from an earlier test is called the *testing effect*, and much research has shown that this effect is robust across many types of materials and kinds of test (see Roediger & Karpicke, 2006a, for an extensive review).

Perhaps more interesting from theoretical and educational standpoints is that taking a test on studied material has frequently been found to produce better learning and retention than additional study of the target material (see, e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006b). The theoretical implication is that the testing effect depends on factors other than just reexposure to or restudy of target material. Even when students receive no feedback on the tests, testing often provides a greater benefit than restudying the material; hence, testing does not simply help memory because of re-presentation; rather, processes involved in retrieval underlie the testing effect (e.g., Carpenter & DeLosh, 2006; Carrier & Pashler, 1992; McDaniel & Masson, 1985). That is, retrieval of information from memory appears to be a potent memory modifier (Bjork, 1975). Testing, then, is not a neutral event in terms of learning, but may itself promote unique processing of target material that improves retention and later test performance.

For educational applications, the clear implication is that testing should be exploited to enhance learning; indeed, testing may be an especially robust technique to assist students in learning. Yet, except in special circumstances (e.g., learning of multiplication tables by flashcards), educators seem not to appreciate the potential of using testing to enhance learning, either in class or as a study strategy to encourage students to use at home. At the level of educational research, the emphasis on testing as a learning device is virtually ignored. Years ago, Glover (1989) published a study entitled "The 'Testing' Phenomenon: Not Gone but Nearly Forgotten," and that state of affairs remains true today. Certainly, in colleges and universities, most (but not all) instructors prefer administering few tests, both because tests are a nuisance to create and grade and because they take up class time.

Of course, one reason that the effects of testing may have been ignored in educational settings is that most experimental research has focused on materials such as word lists and on rather short retention intervals. Little research exists that spans the gap between laboratory studies with such materials and the educational concerns of learning rich, coherent material for long-term retention. To help inform and build the case for the application of the testing effect to education, our research has systematically investigated the testing effect in experiments designed to reflect educationally relevant parameters. In this article,

M. A. McDaniel, mmcdanie@artsci.wustl.edu

we summarize a series of these experiments, which begin in the laboratory and culminate with a parallel experiment situated in a college course.

Before introducing the particular issues on which our studies converge, we should note that testing may have indirect benefits for students and teachers, in addition to the direct benefits on which we focus. In particular, frequent tests encourage students to study more continuously, and the test itself can help expose students to material. If feedback is given, testing can also be used to help shape learning. A recent survey of introductory psychology students at one large state university showed that over two-thirds of the students did not acquire a textbook for the course (Sikorski et al., 2002). Those who did have (or who borrowed) a textbook spent less than 3 h a week reading it. Thus, frequent quizzing may provide the primary exposure of the course content for some students. Testing (quizzing) can also help inform students about what they have and have not learned, thereby potentially aiding subsequent study of course content (see, e.g., Amlund, Kardash, & Kulhavy, 1986). Finally, testing also permits instructors to learn what their students do and do not know and to adjust their teaching strategies accordingly, a process called *formative assessment*. We do not address these and other potential indirect benefits of testing here (see Thomas & McDaniel, 2007, in this issue), although they could significantly amplify the results we do present. Our research concerns the direct positive effects of testing material on later retention of that same material.

Our experiments addressed several overarching issues fundamental to educational application of the testing effect. One central objective was to evaluate the relative benefit of cued-recall tests over recognition (multiple-choice) tests. Studies with laboratory materials (word lists or paired-associate lists) have found that initial recall tests benefit subsequent test performance more than do initial recognition tests (Cooper & Monk, 1976; Darley & Murdock, 1971; Mandler & Rabinowitz, 1981; McDaniel & Masson, 1985; Wenger, Thompson, & Bartling, 1980). The working hypothesis to explain these findings has generally been that tests requiring more effortful responding (like short-answer or essay tests) will generally produce greater benefits on later tests than will tests requiring less effortful responding (e.g., multiple-choice or true/false testing), although other conceptualizations are possible. In our experiments, students studied educationally relevant materials and then took an initial short-answer or multiple-choice test. In companion (control) conditions, students also reread the target facts or were not exposed to them at all after the initial reading. Later, all students were given a final criterial test to determine whether short-answer testing produced the best long-term recall among these conditions.

Another issue addressed in our experiments is the necessity of corrective feedback on the initial tests for obtaining a testing effect. Pashler, Cepeda, Wixted, and Rohrer (2005) reported that final test performance was significantly boosted for items answered correctly on an initial test. However, items missed on the initial test had very little chance of being recalled correctly on a final test unless feedback (in the form of the correct answer) was given (see also McDaniel & Fisher, 1991). Accordingly, we were interested in the degree to which the testing effect varied when the initial test was given with or without feedback.

A third objective was to assess whether testing effects with educational materials reflect processes beyond those evoked by additional study of the recalled material. Accordingly, in all of the present experiments we included additional-exposure conditions in which the target material was presented for additional study instead of being quizzed. Most studies in the educational literature do not clarify this issue, because such extra-study conditions were not generally included (see, e.g., Glover, 1989; Spitzer, 1939). We should note that the use of a control condition in which students are given target materials on which to focus is not externally valid, because normally students would not know what material would be tested, and so could not focus only on that material; rather, they would usually have to read the entire text. Thus, our baseline using focused rereading of facts tested in the other conditions provides a very conservative assessment of the effect of testing.

We briefly report three experiments using educationally relevant materials, each of which is reported in more detail elsewhere. The first is a laboratory study using text materials like those used in courses, the second is another lab study that uses lecture materials and a long retention interval (simulating a classroom situation), and the third is an experiment conducted in a true field setting—a college class.

Laboratory Studies on the Effects of Testing and Feedback

We consider first a pair of experiments by Kang, McDermott, and Roediger (in press), in which the three issues described above were examined with text materials. Specifically, subjects read short articles adapted from *Current Directions in Psychological Science*. Following each article, they either took an initial short-answer or multiple-choice test, or they read key statements from the text. Critically, the information contained in the questions was identical across the testing and reexposure conditions. For example, a short-answer question might be "What is the *hostile media bias*?" The corresponding multiple-choice question (given to other subjects after reading the same article) would be the same, but four possible choices would follow the question. In the "read statements" condition, the subjects would encounter the answer to the question (e.g., *Hostile media bias* refers to the phenomenon in which people on both sides of a controversy perceive the media as being hostile to their group). In one experiment, immediate feedback was provided on the initial tests, whereas in another experiment no feedback was given. Finally, the criterial measure (presented 3 days later) was manipulated; on the final test, some questions were in short-answer format, others in multiple-choice format. Hence, for a short-answer question on the final criterial measure, some subjects would have encountered the same question before (either as a short-answer or multiple-

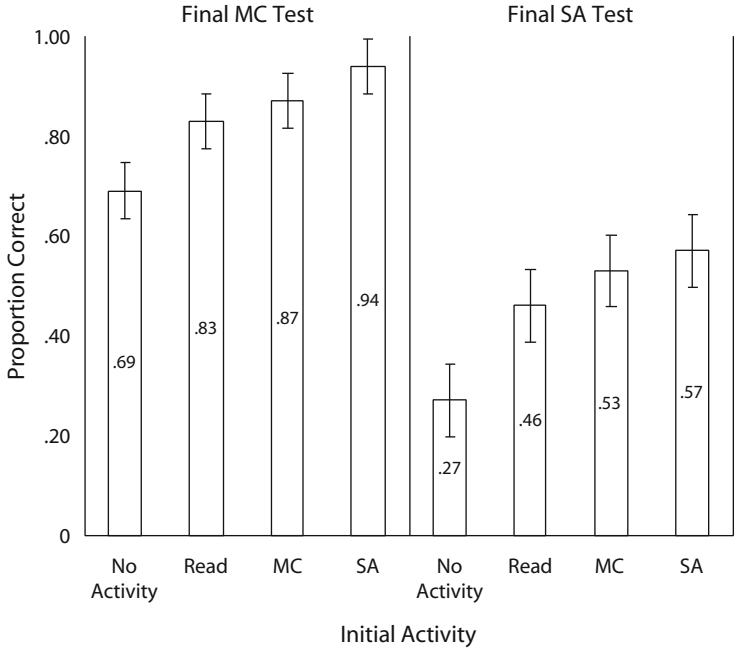

Figure 1. Mean final-test performance as a function of intervening task (from Kang et al., in press, Experiment 2). Error bars show 95% confidence intervals. MC, multiple choice; SA, short answer.

choice question), some would have encountered the answer (in the read statement condition), and some would not have had any initial test or repetition following the initial reading of the passage. This last condition provided a baseline against which the effects of testing and rereading could be measured.

We turn now to the results of the experiment in which immediate feedback was given on the initial test. The results (see Figure 1) show that having taken an initial short-answer test led to the best retention 3 days later, regardless of whether the final criterial measure for that question was in short-answer or multiple-choice format. Taking an initial multiple-choice test also resulted in a testing effect, but a smaller one. That is, the information from multiple-choice test questions (with feedback) was retained better than the information that was not initially tested; this pattern appeared both on the final short-answer and multiple-choice questions. Hence, when immediate feedback is given, initial short-answer tests seem best for promoting long-term retention, regardless of the format of the final test, although a multiple-choice test also confers an advantage (relative to the no-test control condition).

As we noted previously, a separate experiment examined performance using the same design, but without any feedback given on the initial tests. Here, the pattern of results differed somewhat from the one just described. In the absence of initial feedback, initial multiple-choice tests conferred a greater advantage for long-term retention than did initial short-answer tests, and this outcome occurred on both final short-answer and multiple-choice questions. Clearly, the role of feedback is critical in determining the relative mnemonic benefits of initial testing in these experiments. However, we suspect that a crucial variable is the level of performance on the initial test (which in the present study was much higher for multiple-choice than for short-answer tests). The effects of retrieval on an initial test cannot confer benefits under conditions in which performance is low. Kang et al. (in press) suggested that the relatively low level of performance on the initial short-answer test ($M = .54$) versus the initial multiple-choice test (.86) might be the key. In short, the effect of testing can be attenuated when initial test performance is poor (Wenger et al., 1980) and no feedback is given.

Both of the Kang et al. (in press) experiments included a restudy condition that provided a focused rereading of the to-be-tested facts. For example, instead of being tested on the definition of hostile media bias, subjects in that condition reread the definition. As shown in Figure 1, the rereading condition produced better performance than did the no-test condition, and this performance was statistically indistinguishable from the benefit exhibited by the multiple-choice test. However, as noted above, this rereading condition is conservative, since students usually cannot restudy exactly the material that will appear on the test. A more relevant comparison might be to have students reread the entire article once more, as Roediger and Karpicke (2006b) did. Nonetheless, when a short-answer test was required and students received feedback, they later recalled and recognized material better than when they were reexposed to the material in this focused way.

In summary, the Kang et al. (in press) experiments, which used educationally relevant passages in a laboratory setting, showed that the format of the initial and final tests and the presence or absence of feedback all modulate the degree to which initial testing will influence long-term retention. When corrective feedback compen-

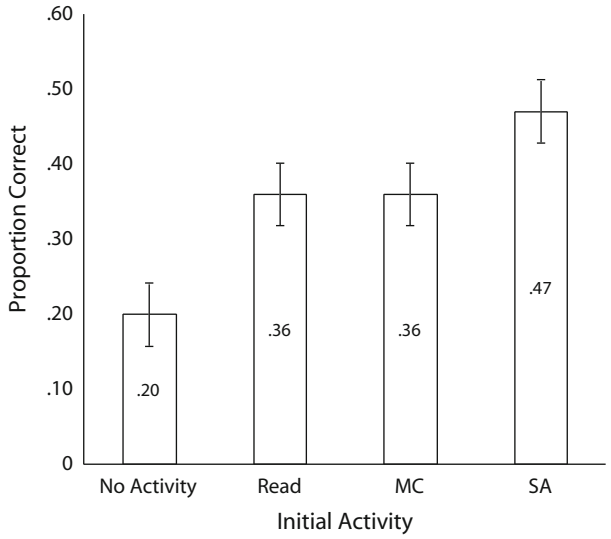

Figure 2. Mean proportions of correct recall for the final short-answer test on art history lectures, as a function of initial postlecture learning condition (from Butler & Roediger, in press). Error bars show 95% confidence intervals. MC, multiple choice; SA, short answer.

sated for poor initial test performance, short-answer tests proved more effective than multiple-choice tests at promoting long-term retention, probably because such tests require greater retrieval effort. We turn now to a different experiment in which we examined the testing effect after students viewed classroom lectures.

The Testing Effect Following Lectures in a Simulated Classroom Setting

Lecture methods are perhaps the most frequently used classroom technique, especially in higher education, yet little research exists on how students learn from lectures and whether testing helps to consolidate learning from them. Butler and Roediger (in press) conducted such an experiment using three art history lectures presented as videos. Students saw a different 30-min lecture each day for three successive days; the lectures focused on different artists, and each was interspersed with slides showing examples of the art. Thirty target items were created for each lecture. After a lecture, students received either a short-answer test, a multiple-choice test, or a focused restudy of the facts. Only 20 of the facts were tested (or restudied), so that the other 10 could be used as a baseline against which to assess the impact of testing or restudying. Furthermore, after the test, students were given feedback on half of the items that were tested (but not the other half) so as to examine the effect of that variable.

Students were given their final, criterial test 30 days after the last lecture. This type of delay simulates the retention interval for a test in a college class. The test was given in short-answer format (e.g., "What aspect of Berthe Morisot's art could be used to date her paintings?"), with students required to write a brief answer. The results are shown in Figure 2, where it can be seen that they agree quite well with the conclusions drawn from the Kang et al. (in press) data in Figure 1. Briefly, previous short-answer tests produced the best performance on the criterial test; the previous multiple-choice and focused restudy conditions produced about equivalent performance; and all three of these conditions produced superior recall relative to the condition in which no repetition occurred after the initial lecture. One surprise was that feedback did not produce any effect in the Butler and Roediger (in press) experiment, and the data in Figure 2 were collapsed across this variable. One possible reason is that performance on the initial test was relatively high, especially for the multiple-choice test (.86); feedback will not be useful if few errors are made. However, this theory cannot explain why feedback did not improve performance on the short-answer test, on which nearly a third of the answers were incorrect (.32). Another possible reason for the ineffectiveness of immediate feedback was that the questions were fairly complex and that students were given just 6 sec to process the feedback, which may have been too little time.

In sum, the Butler and Roediger (in press) experiment confirmed with lecture materials the essential findings of Kang et al. (in press) with text material: A short-answer test produced a greater testing effect than either a multiple-choice test or focused restudying, although both of the latter conditions led to better retention than did studied items that were not repeated. We turn now to see whether these findings can be extended to a college course that students took for credit.

Testing Effects in the Classroom

Testing in a course diverges in important ways from the implementation of testing in laboratory experiments. In experiments, the amount of study time prior to being quizzed is controlled across subjects, whereas in a classroom setting there is variability across students in the amount of studying of the target material. (We label the initial tests here "quizzes" in line with typical terminology for the classroom.) Furthermore, in the experiments above, subjects had no further opportunity to study once they had received the quiz, whereas that is not the case in the classroom. Also, the intervals between exposure to the content and the intervening quiz and between the intervening quiz and the final test were controlled. In contrast, in some classroom situations (such as ad lib quizzing, which can be standard in Web-assisted courses), these delays vary across students. Any of these differences between the class environment and experimental work in a laboratory create uncertainty as to whether the testing effect will generalize to the more variable environment of an actual class.

Accordingly, in conjunction with the research just described, McDaniel, Anderson, Derbish, and Morrisette (in press) implemented an experiment conducted in a college course. The class was a Web-based "Brain and Behavior" course at the University of New Mexico, with weekly reading assignments of approximately 40 pages from the textbook. As in the previous experiments, two kinds of initial tests (short-answer and multiple-choice) and a reading condition featuring the isolated facts targeted by the initial tests were prepared for each weekly assignment. To pro-

vide a no-test control against which to assess the effects of quizzes and reading exposure, each week two sets of 10 facts were identified from the textbook. For each fact in one set, a corresponding fact from the same paragraph was selected for the second set of facts (essentially, facts were yoked across sets). For every week's reading, one set of facts was exposed, either on quizzes or in the reading condition, and the other set of facts was not exposed. The particular facts exposed or not exposed were counterbalanced across students.

The type of exposure (quizzes or rereading) was also manipulated within subjects, such that across 3 weeks of material each student received a 10-item multiple-choice quiz (in one of the weeks), a 10-item short-answer quiz (in another week), and a presentation of 10 facts to read (in the third week). The type of exposure for any particular week's facts was also counterbalanced across students. For example, for the chapter (week) on neurons, a target fact was "All preganglionic axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter." When this fact was in the exposed set, it was presented as a short-answer quiz item for some students, on a multiple-choice quiz for other students, and as a fact to be read for a third group of students. For fact reading, students had to respond to each fact, "I have read the above statement."

The quizzes (or the list of facts to be read) were administered via the Web, and students were allowed to log on and take the assigned quiz (or read the facts) at a time of their choosing. Another important feature of the experiment was that feedback was given following the quiz. The feedback displayed the correct answer, along with a representation of each question and the student's response. For the read condition, in order to equate exposure, the facts were presented again.

After 3 weeks of quizzes, a unit examination was given of the 30 previously quizzed and read items (10 from each week) and the 30 not-previously-tested items. In line with the administration of the course, these unit examinations were always multiple-choice in format. Next, students were given another 3 weeks of quizzes (following an identical design) and then a second unit test, which tested only the material presented in the second 3 weeks of quizzes.

Before reporting the results, we should mention a final feature of this experiment. A critical detail of the laboratory experiments reported above was that the criterial questions were identical to those presented in the initial tests. The same procedure holds for previous work in the educational psychology literature that has shown the testing effect (see, e.g., Glover, 1989; Spitzer, 1939). By contrast, in actual classroom settings, instructors might be loath to administer summative assessments that contain questions identical to those presented on quizzes, and so it was in the present course. Thus, the unit tests focused on the target facts, but the wordings of questions were changed from those used on the quizzes. For instance, in the quiz appeared the stem "All preganglionic axons, whether sympathetic or parasympathetic, release _____ as a neurotransmitter," whereas in the unit test this stem was reworded "All _____ axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter."

The laboratory demonstrations of testing effects summarized here may have produced learning of a particular answer. If so, their testing effects would not necessarily extend to the present classroom context. Alternatively, testing could have stimulated learning of an integrated fact. This more valuable outcome of testing would be implicated if testing effects were to emerge in the present study.

Figure 3 displays the results averaged across the two unit tests. Remarkably, the patterns shadowed those described above for the laboratory experiments. Testing effects were evident, with both short-answer and multiple-choice quizzes augmenting performance on unit examinations relative to when content was not quizzed. Focused reading of the facts, on the other hand, did not boost examination performance. As in the laboratory findings, short-answer quizzing produced significantly higher performance on the unit examinations than did multiple-choice quizzing or focused reading of the target facts. As in Kang et al. (in press), this finding was obtained even though the summative assessment was in multiple-choice format, and therefore matched the multiple-choice quiz format but not the short-answer format.

McDaniel et al. (in press) also administered a final, cumulative assessment at the end of the semester on the material from the Unit 1 and 2 examinations. These units were covered in approximately the middle third of the course, so the final assessment was given approximately 40 days after the Unit 2 examination. This cumulative assessment consisted of 120 multiple-choice items from Units 1 and 2. Thus, half of the items had been tested once (only on the unit examinations) and half twice (once on the quizzes and once on the unit examinations) prior to the cumulative examination. Figure 4 shows that the benefits of short-answer quizzing extended to long retention intervals (as

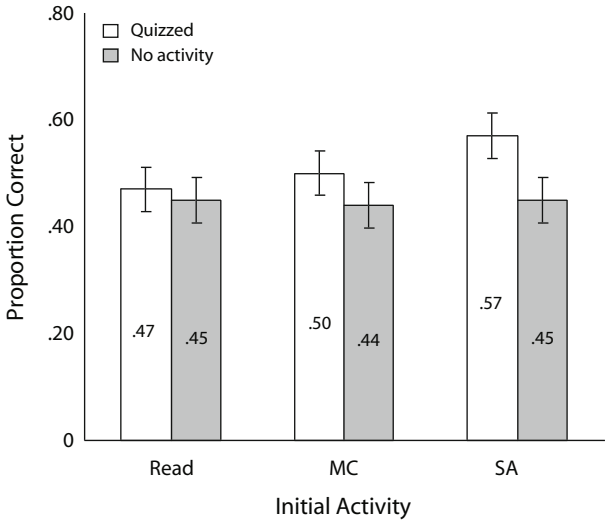

Figure 3. Unit exam performance on quizzed versus nonquizzed items, collapsed across units, in a "Brain and Behavior" course (from McDaniel et al., in press). Error bars show 95% confidence intervals. MC, multiple choice; SA, short answer.

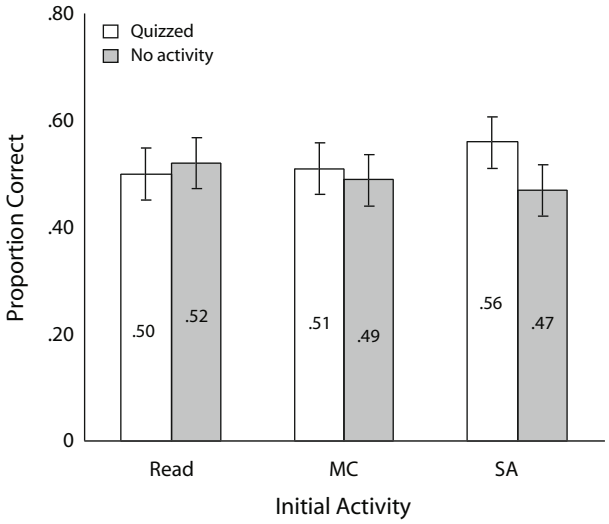

Figure 4. Final-exam performance for quizzed versus nonquizzed items in a "Brain and Behavior" course (from McDaniel et al., in press). Error bars show 95% confidence intervals. MC, multiple choice; SA, short answer.

in Butler & Roediger, in press), even for items that had received testing during the unit examinations prior to the final test. By contrast, the significant benefits of multiple-choice quizzes did not continue to the final cumulative examination; the intervening unit tests alone presumably provided as much benefit (if any; there were no items that had not previously been tested) as was gained from two multiple-choice tests (quiz plus unit test).

These results provide the first experimental demonstration in the context of an actual course that quizzing benefits learning, and that it does so more than focused reading of target facts. Furthermore, this testing effect reflected more than learning of a particular answer. Importantly, quizzes helped students to learn integrated concepts better, since the specific responses on the summative tests differed from those required for the quizzes.

Conclusions

The research presented in this article clearly demonstrates that a robust finding in the experimental memory literature—the potency of tests for augmenting learning and retention—extends to educational materials, educational parameters (long retention intervals), and educational settings (a college course). Furthermore, the fidelity with which these findings match the key patterns in the basic cognitive literature inspires confidence in the fruitfulness of basic research as a means of informing improvements in education. Specifically, several consistent results span a number of designs and environments—from basic laboratory work with word lists, to laboratory experiments with educational materials and significant retention intervals, to an experiment in a college course: (1) Initial short-answer and multiple-choice tests significantly benefit subsequent or final test performance (relative to no tests). (2) Short-answer tests (production or recall of material) benefit subsequent test performance more than do multiple-choice tests (recognition of material), even when the final tests are in multiple-choice format. (3) The benefits of short-answer tests, but not necessarily of multiple-choice tests, significantly exceed those of focused study of the target material. It should be noted that these effects are most prominent when initial tests include corrective feedback.

Given that we obtained these patterns with diverse student populations (a highly selective private university and a less selective public university); diverse materials (short journal articles, videotaped art history lectures, and college textbook content); and retention intervals that varied from several days, to a week, to up to a month, we suggest that testing in education should not be limited to an assessment role. As anticipated by basic experiments from cognitive psychology, testing is a powerful tool to promote learning in educational situations.

AUTHOR NOTE

This research was sponsored by Institute of Education Sciences Grant R305H030339 and a Collaborative Activity Grant from the James S. McDonnell Foundation. We thank Pooja Agarwal and Jane McConnell for assisting with technical aspects of the manuscript, and Andrew Butler and Sean Kang for their comments. Correspondence concerning this article should be addressed to any of the authors at the Department of Psychology, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130-4899.

REFERENCES

Amlund, J. T., Kardash, C. A., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly*, **21**, 49-58.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). New York: Wiley.

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

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

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

Cooper, A. J. R., & Monk, A. (1976). Learning for recall and learning for recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 131-156). New York: Wiley.

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, **91**, 66-73.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, Whole No. 40, 1-104.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, **81**, 392-399.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, **10**, 562-567.

Jones, H. E. (1923/1924). The effects of examination on the performance of learning. *Archives of Psychology*, Whole No. 10, 1-70.

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

Mandler, G., & Rabinowitz, J. C. (1981). Appearance and reality: Does a recognition test really improve subsequent recall and recognition? *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 79-90.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (in press). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, **16**, 192-201.

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

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

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, **1**, 181-210.

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

Sikorski, J. F., Rich, K., Saville, B. K., Buskist, W., Drogan, O., & Davis, S. F. (2002). Student use of introductory texts: Comparative survey findings from two universities. *Teaching of Psychology*, **29**, 312-313.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, **30**, 641-656.

Thomas, A. K., & McDaniel, M. A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding–retrieval interactions. *Psychonomic Bulletin & Review*, **14**, 212-218.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 135-144.