

Research Article

Test-Enhanced Learning

Taking Memory Tests Improves Long-Term Retention

Henry L. Roediger, III, and Jeffrey D. Karpicke

Washington University in St. Louis

ABSTRACT—*Taking a memory test not only assesses what one knows, but also enhances later retention, a phenomenon known as the testing effect. We studied this effect with educationally relevant materials and investigated whether testing facilitates learning only because tests offer an opportunity to restudy material. In two experiments, students studied prose passages and took one or three immediate free-recall tests, without feedback, or restudied the material the same number of times as the students who received tests. Students then took a final retention test 5 min, 2 days, or 1 week later. When the final test was given after 5 min, repeated studying improved recall relative to repeated testing. However, on the delayed tests, prior testing produced substantially greater retention than studying, even though repeated studying increased students' confidence in their ability to remember the material. Testing is a powerful means of improving learning, not just assessing it.*

In educational settings, tests are usually considered devices of assessment. Students take tests in class to assess what they have learned and take standardized tests like the SAT to assess their knowledge and aptitude. In many circumstances, such as university lecture courses, tests are given infrequently (often just two or three times a semester) and are generally perceived as a bother by faculty and students alike. We believe that the neglect of testing in all levels of education is misguided. To state an obvious point, if students know they will be tested regularly (say, once a week, or even every class period), they will study more and will space their studying throughout the semester rather than concentrating it just before exams (see Bangert-Drowns, Kulik, & Kulik, 1991; Leeming, 2002). However, more important for present purposes, testing has a powerful positive effect on future retention. If students are tested on material and successfully recall or recognize it, they will remember it better in

the future than if they had not been tested. This phenomenon, called the testing effect, has been studied sporadically over a long period of time (e.g., Gates, 1917), but is not well known outside cognitive psychology.

Most experiments on the testing effect have been conducted in the verbal learning tradition using word lists (e.g., Hogan & Kintsch, 1971; Izawa, 1967; McDaniel & Masson, 1985; Thompson, Wenger, & Bartling, 1978; Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003) or picture lists (Wheeler & Roediger, 1992) as materials. There have been a few experiments using materials found in educational contexts, beginning with Spitzer (1939; see too Glover, 1989, and McDaniel & Fisher, 1991). However, the title of Glover's article from 17 years ago still sums up the current state of affairs: "The 'testing' phenomenon: Not gone but nearly forgotten."

Our aim in the two experiments reported here was to investigate the testing effect under educationally relevant conditions, using prose materials and free-recall tests without feedback (somewhat akin to essay tests used in education). Most previous research has used tests involving recognition (like multiple-choice tests) or cued recall (like short-answer tests). A second purpose of our experiments was to determine whether testing facilitates learning beyond the benefits of restudying the material. In some testing-effect experiments, a study-test condition is compared with a study-only condition on a delayed retention test. When the subjects in the former condition outperform those in the latter on a final test, one can wonder whether the testing effect is simply due to study-test subjects being reexposed to the material during the test. It is no surprise that students will learn more with two presentations of material rather than one (although some of the word-list experiments cited earlier overcame this problem; see too Carrier & Pashler, 1992; Cull, 2000). To evaluate this restudying explanation of the testing effect, we had students in our control conditions restudy the entire set of material—which should, if anything, bias performance results in favor of this condition, because students who take free-recall tests (without feedback) can only reexperience whatever material they can recall.

Students in our experiments studied short prose passages covering general scientific topics. In Experiment 1, they either

Address correspondence to Henry L. Roediger, III, Department of Psychology, Washington University, Campus Box 1125, One Brookings Dr., St. Louis, MO 63130, e-mail: roediger@arts.wustl.edu.

took a test on the material or studied it again before taking a final retention test 5 min, 2 days, or 1 week later. In Experiment 2, students studied a passage once and took three tests, studied three times and took one test, or studied the passage four times. They then took a final test 5 min or 1 week later. We predicted that performance on immediate retention tests would increase with the number of study opportunities, because massed practice typically produces short-term benefits (e.g., Balota, Duchek, & Paullin, 1989). However, we predicted that taking tests soon after studying would promote superior retention on delayed tests relative to repeatedly studying the material. This outcome would indicate that testing has positive effects on long-term retention above and beyond any effect of re-presentation of the material during the test.

EXPERIMENT 1

Method

Subjects

One hundred twenty Washington University undergraduates, ages 18 to 24, participated in partial fulfillment of course requirements.

Materials

Two prose passages were selected from the reading comprehension section of a test-preparation book for the Test of English as a Foreign Language (TOEFL; Rogers, 2001). Each passage covered a single topic (“The Sun” and “Sea Otters”), and each was divided into 30 idea units for scoring purposes. The passages were 256 and 275 words in length, respectively.

Design

A 2×3 mixed-factorial design was used. Learning condition (restudy vs. test) was manipulated within subjects, and delay of the final test (5 min, 2 days, or 1 week) was manipulated between subjects. The order of learning conditions (restudy or test) and the order of passages (“The Sun” or “Sea Otters”) were counterbalanced across subjects.

Procedure

Subjects were tested during two sessions, in small groups (4 or fewer). They were told that Phase 1 consisted of four 7-min periods and that during any given period they would be asked to study one passage for the first time, restudy one of the passages, or take a recall test over one of the passages. During each study period, subjects read one passage for 7 min. During the test period, subjects were given a test sheet with the title of the to-be-recalled passage printed at the top and were asked to write down as much of the material from the passage as they could remember, without concern for exact wording or correct order. Subjects solved multiplication problems for 2 min between periods and for 5 min after the final period in Phase 1.

Phase 2 occurred after a 5-min, 2-day, or 1-week retention interval. In Phase 2, subjects were asked to recall the passages that they had learned in Phase 1. The recall instructions were identical to those given in Phase 1. Each retention test lasted 10 min, and subjects were instructed by the experimenter to draw a line on their test sheets to mark their place after each 1-min interval during the recall periods (Roediger & Thorpe, 1978). At the end of the experiment, subjects were debriefed and thanked for their participation.

Results and Discussion

Scoring

Subjects’ recall responses were scored by giving 1 point for each correctly recalled idea unit (out of 30). Initially, 40 recall tests were scored by two raters, and the Pearson product-moment correlation (r) between their scores was .95. Given the high interrater reliability, the remaining recall tests were scored by one rater.

Initial Test

On the initial 7-min test, subjects recalled on average 20.9 idea units, or approximately 70% of the passage. No differences were observed for the two passages or for the different counterbalancing orders.

Final Test

The mean proportion of idea units recalled on the final tests after the three retention intervals is shown in Figure 1. The cumulative recall data showed that subjects had exhausted their knowledge by the end of the retention interval and are not reported here. After 5 min, subjects who had studied the passage

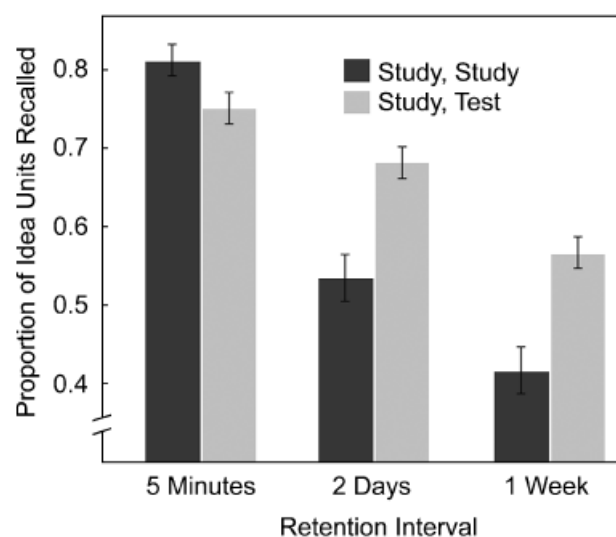


Fig. 1. Mean proportion of idea units recalled on the final test after a 5-min, 2-day, or 1-week retention interval as a function of learning condition (additional studying vs. initial testing) in Experiment 1. Error bars represent standard errors of the means.

twice recalled more than subjects who had studied once and taken a recall test. However, this pattern of results was reversed on the delayed tests 2 days and 1 week later. On these tests of long-term retention, subjects who had taken an initial test recalled more than subjects who had only studied the passages.

The results were submitted to a 2×3 analysis of variance (ANOVA), with learning condition (restudying or testing) and retention interval (5 min, 2 days, or 1 week) as independent variables. This analysis revealed a main effect of testing versus restudying, $F(1, 117) = 36.39$, $\eta_p^2 = .24$, which indicated that, overall, initial testing produced better final recall than additional studying. Also, the analysis revealed a main effect of retention interval, $F(2, 117) = 50.34$, $\eta_p^2 = .46$, which indicated that forgetting occurred as the retention interval grew longer. However, these main effects were qualified by a significant Learning Condition \times Retention Interval interaction, $F(2, 117) = 32.10$, $\eta_p^2 = .35$, indicating that restudying produced better performance on the 5-min test, but testing produced better performance on the 2-day and 1-week tests.

Post hoc analyses confirmed that on the 5-min retention tests, restudying produced better recall than testing (81% vs. 75%), $t(39) = 3.22$, $d = 0.52$. However, the opposite pattern of results was observed on the delayed retention tests. After 2 days, the initially tested group recalled more than the additional-study group (68% vs. 54%), $t(39) = 6.97$, $d = 0.95$. The benefits of initial testing were also observed after 1 week: The tested group recalled 56% of the material, whereas the restudy group recalled only 42%, $t(39) = 6.41$, $d = 0.83$. Figure 1 depicts another interesting finding: The initially tested group recalled as much on the 1-week retention test as the additional-study group did after only 2 days (the initially tested group actually recalled slightly more). This surprising result indicates that taking an initial recall test prevented forgetting of information for an additional 5 days relative to repeated study.

Experiment 1 demonstrated that after an initial study episode, additional studying or testing had different effects on immediate and delayed final tests: Relative to testing, additional studying aided performance on immediate retention tests; in contrast, prior testing improved performance on delayed tests. The crossover interaction observed in Figure 1 is all the more impressive considering that no feedback was given on the tests. The testing effect on delayed retention tests is not simply due to reexposure to studied material during tests, but rather is due to some other process that has positive effects on retention. We consider candidate processes in the General Discussion.

EXPERIMENT 2

In Experiment 2, we investigated the effects of repeated studying and repeated testing on retention, in part to replicate and extend the results of Experiment 1, but more to ask about effects of repeated testing. We were interested in the effects of repeated testing because most testing-effect experiments compare per-

formance on final tests after subjects have or have not taken a single test earlier, as we did in Experiment 1. However, Wheeler and Roediger (1992) showed that taking three tests immediately after studying a list of pictures greatly improved retention on a final test relative to taking a single test or no test. In Experiment 2, we compared three conditions: Subjects studied a passage four times (and took no tests), studied it three times and took one test, or studied it once and took three tests. They then took a final test either 5 min or 1 week later. In addition to examining effects of repeated versus single tests, we made a few procedural changes to obtain estimates of how many times students in the various conditions actually read each passage. We also included a brief questionnaire after the initial learning session, asking subjects to rate how interesting and readable they found the passage and, more important, how well they thought they would remember it on a test 1 week later. We were particularly interested in subjects' predictions of how well they would remember the passage, because such judgments are not always correlated with actual performance (see Bjork, 1994; Koriat, Bjork, Sheffer, & Bar, 2004).

Method

Subjects and Materials

One hundred eighty Washington University undergraduates, ages 18 to 24, participated in partial fulfillment of course requirements. The passages used in Experiment 1 were used again and were counterbalanced across conditions.

Design

A 3×2 between-subjects design was used. Subjects learned one of the two prose passages under one of three conditions (S = study, T = test): repeated study (SSSS), single test (SSST), or repeated test (STTT). Ninety subjects were given a final recall test following a 5-min retention interval, and 90 took a final test after 1 week. Thirty subjects were assigned to each of the six between-subjects conditions.

Procedure

The procedure used in Experiment 2 was similar to that used in Experiment 1. Subjects were again tested during two sessions, in small groups (4 or fewer). In Phase 1, they were told that they would be learning one passage during four consecutive periods. Subjects in the SSSS condition read the passage during four 5-min study periods; subjects in the SSST condition studied the passage during three periods and then took one recall test; those in the STTT condition studied the passage during one period and then took three consecutive recall tests. Students in the multiple-test condition were instructed to try hard to recall the entire passage on each successive test.

During study periods, subjects had 5 min to study the passage, and they recorded the number of times they read the entire passage by making tally marks on a separate sheet. During test

periods, subjects were given a blank sheet and were asked to recall as much of the material from the passage as they could remember, without concern for exact wording or correct order. Each test lasted 10 min, and subjects were instructed to draw a line on their test sheets to mark their place after each 1-min interval. Subjects solved multiplication problems for 2 min between periods and for 5 min after the final period in Phase 1.

At the end of Phase 1, subjects were given a questionnaire asking them to answer three questions using a 7-point scale. They indicated how interesting they thought the passage was (1 = *very boring*, 7 = *very interesting*), how readable they thought it was (1 = *very easy to read*, 7 = *very difficult to read*), and how well they thought they would remember the passage in 1 week (1 = *not very well*, 7 = *very well*). After completing the questionnaire, subjects in the 5-min retention-interval condition took the final recall test, and subjects in the 1-week condition were excused, returning for the final test 1 week later. The final recall test (Phase 2) was identical to the initial recall tests.

Results and Discussion

Readings of the Passage

The mean number of times subjects were able to read through the passage during each study period is presented in Table 1. No differences in these reading scores were observed for the two passages or for the 5-min and 1-week retention-interval groups. Across all conditions, subjects were able to read the entire passage approximately 3.5 times during a 5-min study period. The number of times subjects in the SSSS and SSST conditions read the passage increased slightly across consecutive study periods, $F(3, 177) = 1.62, \eta_p^2 = .03$, and $F(2, 118) = 4.99, \eta_p^2 = .08$, respectively. The reading scores in Table 1 simply illustrate that subjects read the passage many more times in the SSSS ($M = 14.2$) and SSST ($M = 10.3$) conditions than in the STTT ($M = 3.4$) condition.

Initial Tests

Subjects in the STTT condition recalled 20.9, 21.2, and 21.1 idea units on each of the three initial recall tests, respectively, or about 70% of the passage in each case. No differences on the initial tests were observed for the two passages or for the 5-min

TABLE 1
Mean Number of Times Subjects Were Able to Read the Entire Passage During the 5-Min Study Periods in Experiment 2

Condition	Study period				Sum
	1	2	3	4	
SSSS	3.4	3.5	3.6	3.7	14.2
SSST	3.2	3.5	3.6	—	10.3
STTT	3.4	—	—	—	3.4

Note. Condition labels indicate the order of study (S) and test (T) periods.

TABLE 2
Mean Ratings on the Questionnaire Given After the Initial Learning Session in Experiment 2

Condition	Rating		
	Interesting	Readable	Remember
SSSS	3.8	2.5	4.8
SSST	4.1	2.5	4.2
STTT	4.6	2.8	4.0

Note. Condition labels indicate the order of study (S) and test (T) periods. Subjects rated how interesting the passage was (1 = *very boring*, 7 = *very interesting*), how readable the passage was (1 = *very easy to read*, 7 = *very difficult to read*), and how well they believed they would remember the passage in 1 week (1 = *not very well*, 7 = *very well*).

and 1-week retention-interval groups. Measures of cumulative recall indicated that asymptotic levels of recall had been reached by the end of each test period. Subjects in the SSST condition recalled 23.1 idea units (77% of the passage) on their initial recall test. This was reliably greater recall than on the third test in the STTT condition, $t(118) = 3.17, d = 0.58$.

Questionnaire

The mean ratings on the questionnaire given at the end of Phase 1 are displayed in Table 2. No differences in the questionnaire ratings were observed for the two passages or for the 5-min and 1-week retention-interval groups. Subjects in the SSSS condition rated the passage as less interesting than subjects in the SSST or STTT condition, $F(2, 177) = 3.88, \eta^2 = .04$, perhaps because of increased boredom with repeated readings. More interestingly, subjects in the SSSS condition were more confident that they would remember the passage in 1 week than were subjects in the SSST or STTT condition, $F(2, 177) = 6.09, \eta^2 = .06$. Post hoc analyses revealed that subjects in the SSSS condition predicted that they would remember the passage better than subjects in the SSST condition, $t(118) = 2.95, d = 0.54$, and subjects in the STTT condition, $t(118) = 3.35, d = 0.61$, but the latter two groups did not differ significantly in their predictions. The three groups did not differ in how they rated the readability of the passages ($F < 1$).

Final Tests

The critical data are the mean proportions of idea units recalled on the final tests 5 min or 1 week later, displayed in Figure 2. The pattern of final test scores replicates the pattern of results found in Experiment 1. On the 5-min test, recall was correlated with repeated studying: The SSSS group recalled more than the SSST group (83% vs. 78%), who in turn recalled more than the STTT group (71%). However, on the 1-week test, recall was correlated with the number of tests given earlier: The STTT group recalled more than the SSST group (61% vs. 56%), who in turn recalled more than the SSSS group (40%).

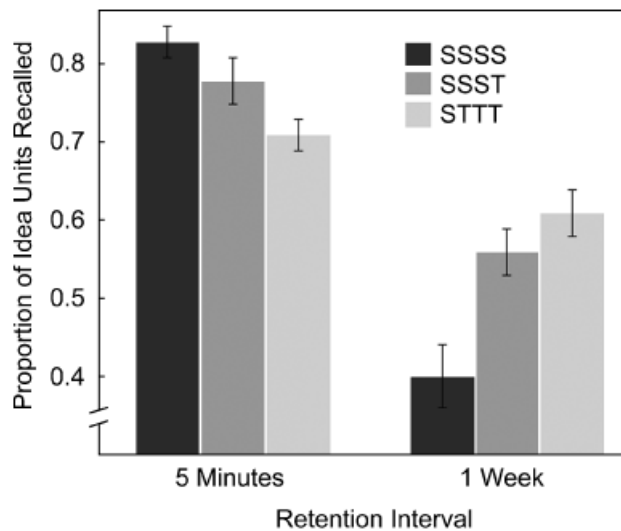


Fig. 2. Mean proportion of idea units recalled on the final test after a 5-min or 1-week retention interval as a function of learning condition (SSSS, SSST, or STTT) in Experiment 2. The labels for the learning conditions indicate the order of study (S) and test (T) periods. Error bars represent standard errors of the means.

The results in Figure 2 were submitted to a 2×3 ANOVA, with retention interval (5 min or 1 week) and learning condition (SSSS, SSST, or STTT) as independent variables. This analysis revealed a main effect of retention interval, $F(1, 174) = 122.53$, $\eta_p^2 = .41$, indicating that forgetting occurred. The effect of learning condition was only marginally significant, $F(2, 174) = 2.32$, $\eta_p^2 = .03$. However, these effects were qualified by a significant Learning Condition \times Retention Interval interaction, $F(2, 174) = 18.48$, $\eta_p^2 = .18$. This interaction indicates that repeated studying produced short-term benefits, whereas repeated testing produced greater benefits on the delayed test.

Post hoc analyses confirmed these observations. On the 5-min retention test, subjects in the SSSS condition recalled more than subjects in the STTT condition, $t(58) = 4.70$, $d = 1.22$, as did subjects in the SSST condition, $t(58) = 2.24$, $d = 0.59$. However, this pattern was reversed on the 1-week retention test. Subjects in the STTT condition recalled more than subjects in the SSST condition, though this difference was marginal, $t(58) = 1.21$, $d = 0.31$. Subjects in the STTT condition also recalled significantly more than subjects in the SSSS condition, $t(58) = 4.78$, $d = 1.26$, as did subjects in the SSST condition, $t(58) = 3.21$, $d = 0.82$.

These data, involving absolute measures of forgetting, show greater forgetting in the pure-study condition than in the testing conditions. An alternative approach to examining forgetting is to use a proportional measure: (initial recall – final recall)/initial recall. In some studies, this alternative has led to different conclusions about rates of forgetting (Loftus, 1985). Proportional measures of forgetting are presented in Figure 3, in which it is obvious that subjects in the SSSS condition forgot far more (52%) than subjects in the SSST condition (28%) and than

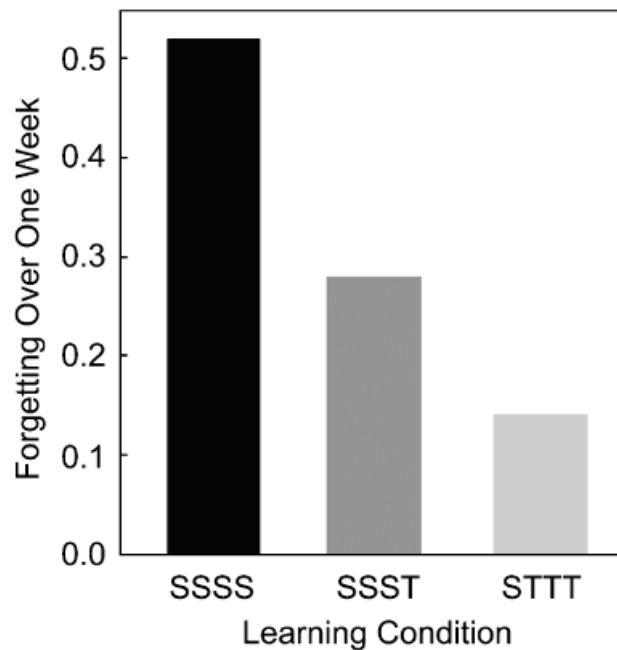


Fig. 3. Forgetting over 1 week as a function of learning condition (SSSS, SSST, or STTT) in Experiment 2. The labels for the learning conditions indicate the order of study (S) and test (T) periods.

subjects in the STTT condition (14%). Subjects in the SSST condition forgot more than subjects in the STTT condition (28% vs. 14%). Thus, the proportional-forgetting analyses confirm those using the raw data and clearly demonstrate the powerful effect of repeated testing in preventing forgetting (cf. Wheeler & Roediger, 1992).

GENERAL DISCUSSION

Both experiments showed the same pattern: Immediate testing after reading a prose passage promoted better long-term retention than repeatedly studying the passage. This outcome occurred even though the tests included no feedback. Clearly, the testing effect is not simply a result of students gaining reexposure to the material during testing, because restudying allowed students to reexperience 100% of the material but produced poor long-term retention (see too Wheeler et al., 2003). The positive effects of testing were dramatic: In Experiment 2, students in the repeated-testing condition recalled much more after a week than did students in the repeated-study condition (61% vs. 40%), even though students in the former condition read the passage only 3.4 times and those in the latter condition read it 14.2 times. Testing has a powerful effect on long-term retention.

The situation was different for tests taken shortly after learning: Repeated studying improved performance relative to repeated testing on final tests given after a 5-min retention interval, but the effect reversed on delayed tests. This pattern of results is analogous to the finding in the spacing-effect literature that massed presentation improves performance on immediate

tests, whereas spaced presentation leads to better performance on delayed tests (Balota et al., 1989; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963). That is, in both cases, massed study leads to a short-term benefit, but the other manipulation (testing or spaced studying) has a greater effect on long-term retention. Both outcomes may reflect the role of desirable difficulties in promoting long-term retention (Bjork, 1994), as discussed later. This outcome on the immediate tests in the present experiments reveals just how powerful the testing effect is: Despite the benefits of repeated study shortly after learning, repeated testing produces strong positive effects on a delayed test.

Several overlapping theoretical approaches are useful in understanding our results. The findings are consistent with theories of transfer-appropriate processing that emphasize the compatibility between the operations engaged in during learning and testing phases (Morris, Bransford, & Franks, 1977; Roediger, 1990). The ability to remember a prose passage on a free-recall test a week after learning it is enhanced by practicing exactly this skill during learning. Practicing the skills during learning that are needed during retrieval generally enhances retention on both explicit and implicit memory tests (Roediger, Gallo, & Geraci, 2002). Although restudying the passages exposed students to the entire set of information, testing permitted practice of the skill required on future tests and hence enhanced performance after a delay.

McDaniel and his colleagues (McDaniel & Fisher, 1991; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985) have argued that testing enhances learning by producing elaboration of existing memory traces and their cue-target relationships, and Bjork (1975, 1988) has suggested that testing operates by multiplying the number of “retrieval routes” to stored events. Bjork (1994, 1999) has also emphasized the need to introduce desirable difficulties into training and educational settings. Many study conditions and strategies that produce rapid learning and short-term benefits lead to poor long-term performance. Our results show that testing versus studying is another case in point: Testing clearly introduced a desirable difficulty during learning.

Relative to testing, repeated studying inflated students’ confidence in their ability to remember the passages in the future, even though repeated-study subjects actually showed much poorer retention on delayed tests. Repeated studying is a strategy that students frequently report using and is often recommended to students by teachers (see Rawson & Kintsch, 2005, for discussion). Students may prefer repeated studying because it produces short-term benefits, and students often use ineffective learning strategies because they base their predictions of future performance on what produces rapid short-term gains. Although students in the repeated-study condition predicted they would perform very well a week later (relative to those in the other conditions), they actually performed the worst.

Free-recall testing even without feedback had large positive effects on retention in our experiments. Pashler and his col-

leagues (Pashler, Cepeda, Wixted, & Rohrer, 2005; Pashler, Zarow, & Triplett, 2003) have examined testing effects with feedback in paired-associates paradigms and also reported positive effects. Testing with feedback may improve performance even beyond the levels observed in the current research (McDermott, Kang, & Roediger, 2005). Judicious use of testing may improve performance in educational settings at all levels from elementary through university education, at least in fact-based courses (Roediger & Karpicke, 2006). Frequent testing leads students to space their study efforts, permits them and their instructors to assess their knowledge on an ongoing basis, and—most important for present purposes—serves as a powerful mnemonic aid for future retention. The boundary conditions for the testing effect are not yet known, but we suspect that tests will produce strong effects when they occur relatively soon after learning and permit relatively high levels of performance (Balota, Duchek, Sergent-Marshall, & Roediger, in press; Landauer & Bjork, 1978; Logan & Balota, 2005; Spitzer, 1939). We believe the time is ripe for a thorough examination of the mnemonic benefits of testing and its potentially important consequences for improving educational practice.

Acknowledgments—This research was supported by a grant from the Institute of Educational Sciences and by a Collaborative Activity Grant from the James S. McDonnell Foundation. We thank Jane McConnell for her help and John Wixted for comments.

REFERENCES

- Balota, D.A., Duchek, J.M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging, 4*, 3–9.
- Balota, D.A., Duchek, J.M., Sergent-Marshall, S., & Roediger, H.L., III. (in press). Does expanded retrieval produce benefits over equal interval spacing? Explorations in healthy aging and early stage Alzheimer’s disease. *Psychology and Aging*.
- Bangert-Drowns, R.L., Kulik, J.A., & Kulik, C.L.C. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 85*, 89–99.
- 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). Hillsdale, NJ: Erlbaum.
- Bjork, R.A. (1988). Retrieval practice and the maintenance of knowledge. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 396–401). New York: Wiley.
- Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bjork, R.A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.
- Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235.
- Gates, A.I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*(40).
- 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 and Verbal Behavior*, *10*, 562–567.
- Izawa, C. (1967). Function of test trials in paired-associate learning. *Journal of Experimental Psychology*, *75*, 194–209.
- Koriat, A., Bjork, R.A., Sheffer, L., & Bar, S.K. (2004). Predicting one’s own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656.
- Landauer, T.K., & Bjork, R.A. (1978). Optimum rehearsal patterns and name learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
- Leeming, F.C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*, 210–212.
- Loftus, G. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 397–406.
- Logan, J.M., & Balota, D.A. (2005). *Spaced and expanded retrieval effects in younger and older adults*. Manuscript submitted for publication.
- McDaniel, M.A., & Fisher, R.P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192–201.
- McDaniel, M.A., Kowitz, M.D., & Dunay, P.K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, *17*, 423–434.
- McDaniel, M.A., & Masson, M.E.J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371–385.
- McDermott, K.B., Kang, S., & Roediger, H.L., III. (2005, January). *Test format and its modulation of the testing effect*. Paper presented at the biennial meeting of the Society for Applied Research in Memory and Cognition, Wellington, New Zealand.
- Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.
- 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.
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1051–1057.
- Peterson, L.R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing of presentations on retention of paired-associates over short intervals. *Journal of Experimental Psychology*, *66*, 206–209.
- Rawson, K.A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*, 70–80.
- Roediger, H.L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043–1056.
- Roediger, H.L., III, Gallo, D.A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory*, *10*, 319–332.
- Roediger, H.L., III, & Karpicke, J.D. (2006). *The power of testing memory: Implications for educational practice*. Unpublished manuscript, Washington University in St. Louis.
- Roediger, H.L., III, & Thorpe, L.A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition*, *6*, 296–305.
- Rogers, B. (2001). *TOEFL CBT Success*. Princeton, NJ: Peterson’s.
- Spitzer, H.F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.
- Thompson, C.P., Wenger, S.K., & Bartling, C.A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210–221.
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175–184.
- Wheeler, M.A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580.
- Wheeler, M.A., & Roediger, H.L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard’s (1913) and Bartlett’s (1932) results. *Psychological Science*, *3*, 240–245.

(RECEIVED 2/4/05; ACCEPTED 3/24/05;
FINAL MATERIALS RECEIVED 4/13/05)