

# Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention

BENJAMIN C. STORM

*University of Illinois, Chicago, Illinois*

AND

ROBERT A. BJORK AND JENNIFER C. STORM

*University of California, Los Angeles, California*

Retrieving information from memory makes that information more recallable in the future than it otherwise would have been. Optimizing retrieval practice has been assumed, on the basis of evidence and arguments tracing back to Landauer and Bjork (1978), to require an expanding-interval schedule of successive retrievals, but recent findings suggest that expanding retrieval practice may be inferior to uniform-interval retrieval practice when memory is tested after a long retention interval. We report three experiments in which participants read educational passages and were then repeatedly tested, without feedback, after an expanding or uniform sequence of intervals. On a test 1 week later, recall was enhanced by the expanding schedule, but only when the task between successive retrievals was highly interfering with memory for the passage. These results suggest that the extent to which learners benefit from expanding retrieval practice depends on the degree to which the to-be-learned information is vulnerable to forgetting.

Tests are commonly used in educational settings as a means of assessing the state of a student's knowledge. Research has shown, however, that tests do much more than measure learning; they also enhance learning (e.g., Bjork, 1975, 1988; Carrier & Pashler, 1992; Glover, 1989; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Roediger & Karpicke, 2006b; Spitzer, 1939; Tulving, 1967; Wheeler & Roediger, 1992). Not only does information that has been tested become more recallable in the future than it would have been otherwise, that information, if retrieved, becomes more recallable than if such a test was replaced by an additional study opportunity. Testing as pedagogy, therefore, versus as assessment, seems to have great potential for application in training and educational contexts (see, e.g., Bjork, 1994a; Roediger & Karpicke, 2006a).

An important aspect of tests as learning events is that the deeper, more difficult, and more complex retrieval is, the more powerful that retrieval will be in facilitating successful retrievals in the future (e.g., Bjork, 1975; Whitten & Bjork, 1977). Tests that require learners to engage in deep and elaborative retrieval processes are likely to be highly effective; tests that require only superficial processing—such as, in the limit, retrieving very recent information from short-term memory—are not. One simple way of making tests more difficult—and therefore inducing a deeper level of processing—is by delaying the time between learning and test. When tests are given im-

mediately, learners are able to access information from memory in a way that affords little or no benefit above and beyond simply having such information re-presented to them or even beyond not having the information tested or re-presented. When tests are delayed, however, and the to-be-tested information has become less accessible, learners are forced to engage in the type of processing that promotes learning and long-term retention (e.g., Cull, 2000; Glover, 1989; Jacoby, 1978; Modigliani, 1976; Roediger & Karpicke, 2006b; Whitten & Bjork, 1977). Said differently, delayed tests constitute better practice for later recall because they exercise more of the processes needed to succeed on a later test (for an embellishment of that argument, see Bjork, 1988).

With the benefit of a delayed test, however, also comes a potential danger. In order for an item to profit from being tested, the learner must be able to successfully retrieve that item from memory, and the likelihood of doing so decreases with the delay between learning and test. Thus, there is a dilemma: If the delay between learning and test is short, retrieval is likely to succeed but to be ineffectual; if the delay is long, retrieval is unlikely to succeed and, hence, also to be ineffectual. One potential way of dealing with this dilemma is by implementing an expanding schedule of tests. In order to ensure successful retrieval, initial tests should be relatively immediate, and then, as the to-be-learned information gains strength in memory,

---

B. C. Storm, [bstorm@uic.edu](mailto:bstorm@uic.edu)

---

the interval between successive tests can be systematically increased. By employing such an expanding schedule of tests, learners may be able to benefit from the positive effects of delayed tests while not being harmed by recall failures.

Landauer and Bjork (1978) tested this idea by having participants learn and be tested on the names of fictitious people under various testing schedules. The procedure involved having the participants study names paired with faces (or last names paired with first names) and then being repeatedly tested on those items via cued recall. Whereas some items were tested with an expanding schedule of retrieval practice (e.g., 1 then 4 then 10 interpolated study or test trials on other names before each of three successive tests on a given name), other items were tested with a uniform schedule of retrieval practice (e.g., 5 then 5 then 5 interpolated trials). Note that the total amount of spacing (15 intervening trials) was the same in both the expanding and uniform conditions. When the participants were tested after a 30-min retention interval, recall for items tested via the expanding schedule was significantly better than recall for items tested via the uniform schedule.

In light of these results, many researchers have argued that expanding retrieval practice has great potential to facilitate learning (e.g., Cull, Shaughnessy, & Zechmeister, 1996; Dempster & Perkins, 1993; Rea & Modigliani, 1985). As was emphasized by Bjork (1988), expanding retrieval practice also has two intrinsic advantages, one of which is a low failure rate, which can be important for some populations, such as children and patients. The second advantage, relative to demanding mnemonic techniques, such as interactive imagery, is that expanding retrieval practice can ensure that a particular piece of information is maintained in memory without being lost during the effort to generate an image or story. It is not surprising, therefore, that expanding retrieval practice has been shown to improve memory retention in amnesic patients (Schacter, Rich, & Stamp, 1985) and individuals with dementia (Camp, 2006) and that it has become an important tool in cognitive rehabilitation (Wilson, Baddeley, Evans, & Sheil, 1994).

Recently, however, several studies have failed to replicate Landauer and Bjork's (1978) findings. In some instances, there has been a failure to find an advantage of expanding retrieval practice (e.g., Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Carpenter & DeLosh, 2005), and in other studies, uniform retrieval practice has actually led to significantly better performance than has expanding retrieval practice (e.g., Cull, 2000; Karpicke & Roediger, 2007; Logan & Balota, 2008) when memory was tested after a long retention interval (e.g., days). These and other findings have called into question whether expanding retrieval practice is the most effective practice schedule for learning and for long-term recall (for a recent review, see Balota, Duchek, & Logan, 2007).

In a study particularly relevant to the present investigation, Karpicke and Roediger (2007) had participants learn a series of Graduate Record Exam vocabulary pairs and then tested them on those pairs via various retrieval

practice schedules. Whereas some participants were given a final test following a 10-min retention interval, others were given a final test following a 2-day retention interval. Although Karpicke and Roediger replicated Landauer and Bjork's (1978) expanding retrieval practice effect after 10 min, they found the opposite pattern of results after 2 days. More specifically, whereas the expanding schedule of tests improved recall in the short term, the uniform schedule of tests improved recall in the long term. In explaining their results, Karpicke and Roediger argued that the first retrieval is most important in terms of facilitating long-term retention and that the scheduling of the tests that follow that initial retrieval matters to a much lesser extent.

If expanding retrieval practice does not enhance long-term retention—and in some cases even leads to worse recall (relative to uniform retrieval practice)—it is clearly less than ideal for application within educational contexts. We believe, however, that such a conclusion is unwarranted, or at least is too general. In Landauer and Bjork's (1978) study, there was substantial forgetting of the to-be-learned materials across trials, meaning that considerably more forgetting took place before the initial test in the uniform condition than in the expanding condition. Landauer and Bjork chose to examine memory for people's names, in fact, not only because it is a real-world case in which there is often only a single presentation of something to be remembered, but also because—as we are all painfully aware—names are very quickly forgotten without retrieval practice. In Karpicke and Roediger's (2007) study, however, performance on the initial test was only moderately better in the expanding condition than it was in the uniform condition (.78 versus .73 in their Experiment 1). To the extent that expanding retrieval practice improves learning by increasing the number of items that can be successfully retrieved on subsequent tests, the advantages of expanding practice may be lessened, or reversed, for materials that are forgotten more slowly.

In the experiments reported in the present article, we examined whether the optimal scheduling of retrieval practice is dependent on how susceptible the to-be-learned material is to forgetting. In the first experiment, participants studied an educational passage about Antarctica and were then instructed to free recall information about Antarctica four times, via either a uniform- or expanding-interval schedule. The participants in Experiment 1A were given 5 min to study the passage, whereas the participants in Experiment 1B were given 1 min to study the passage. We predicted that an expanding schedule of retrieval practice would prove more effective than a uniform schedule of retrieval practice to the extent that the participants' memory for the Antarctica passage was vulnerable to forgetting during the interpolated intervals.

## EXPERIMENTS 1A AND 1B

### Method

Half of the participants in Experiments 1A and 1B were asked to recall the passage on Antarctica after 0, 3, 7, and 18 intervening

minutes of interpolated activity, and the remaining participants were asked to recall the passage after 7, 7, 7, and 7 intervening minutes of interpolated activity. Memory was assessed on a surprise final recall task administered after a 1-week retention interval.

**Participants.** A total of 88 undergraduate students (26 male and 62 female) from the University of California, Los Angeles, participated for partial credit in an introductory psychology course. The students were on average 19.5 years old. The first 50 participants were given 5 min to study the passage (Experiment 1A), whereas the final 38 participants were given 1 min to study the passage (Experiment 1B). All of the participants were drawn from the same pool of individuals.

**Materials.** An educational passage about Antarctica was created for the participants to study. The 203-word passage fit entirely on a single 8 × 11 in. piece of paper and was divided into four separate paragraphs. Each paragraph discussed a different subset of information about Antarctica (i.e., geography, climate, people, and location). A total of 15 critical facts were selected throughout the passage for the purpose of scoring (e.g., *The coldest temperature ever recorded at the South Pole is -88° C*). During the intertrial intervals, the participants were asked to read a passage about the constitutional beginnings of American government and politics. The interpolated passage was 34 pages long and was taken directly from a university-level textbook.

**Procedure.** On arrival, each participant was seated at a desk, informed as to the nature of the experiment, and randomly assigned to the expanding or uniform condition. All of the participants were then given the passage about Antarctica and told to study it. The participants in Experiment 1A were given 5 min to study the passage, whereas the participants in Experiment 1B were given 1 min to study the passage. Each participant was then tested four times in accordance with either the expanding or uniform condition. During each of the tests, the participants were given 4 min to write down as much information from the passage as possible.

The expanding schedule consisted of an immediate test followed by three additional tests after 3-, 7-, and 18-min periods of studying the interpolated passage on American government and politics. The uniform schedule consisted of a first test after 7 min of reading the interpolated passage, plus three additional tests after 7, 7, and 7 min of additional study of the interpolated passage. The total time of the interpolated reading was, therefore, 28 min in both conditions. The participants were warned that they might be tested on the interpolated material at the end of the experiment.

After completing the fourth free recall test, the participants were informed that the first phase of the experiment was complete and that they were to return 1 week later to complete the experiment. When they returned, each participant was given two final tests: free recall and cued recall. The free recall test was identical to the four tests that they had taken during the first phase of the experiment. The cued recall test consisted of a series of fill-in-the-blank questions for which the participants had to fill in a missing keyword. There was a total of 15 fill-in-the-blank questions, each representing a different

critical fact from the passage. The order of the two tests was always the same, with the free recall test preceding the cued recall test.

**Results**

Performance on the free recall tests was measured by recording the proportion of critical facts that each participant successfully recalled about Antarctica.

**Learning-phase recall.** The proportion of critical facts recalled correctly on the initial four tests is shown in Table 1 as a function of test number, initial study time, and practice schedule. These data were subjected to a 4 (Test 1 vs. Test 2 vs. Test 3 vs. Test 4) × 2 (expanding vs. uniform) × 2 (1 min vs. 5 min) mixed design ANOVA, with test schedule and study time serving as between-subjects variables. Note that this was a between-experiments ANOVA. We found it appropriate to analyze the data this way, given that the participants were recruited from the same pool of individuals and that the only difference between Experiments 1A and 1B was the amount of time that the participants were given to study the passage. Furthermore, as can be seen in Table 1, the general pattern of results was very similar in the two experiments. Although the participants given 1 min to study the passage performed significantly worse than the participants given 5 min to study the passage [ $F(1,84) = 56.29, MS_e = 5.12, p < .001$ ], the effect of study time did not interact with test number or practice schedule. As such, the analyses reported below collapse across participants in the two experiments.

Collapsing across the four initial tests, we found no evidence of a significant difference in performance between the expanding ( $M = .55, SE = .02$ ) and uniform ( $M = .52, SE = .02$ ) conditions [ $F(1,84) < 1, p > .05$ ], which is somewhat surprising, given that the expanding practice schedule was designed to prevent the participants from forgetting information about the passage before taking the initial test. An interaction emerged between test number and practice schedule [ $F(3,252) = 3.23, MS_e = 0.01, p < .05$ ]. For the participants in the expanding condition, recall performance appeared to stay approximately the same, or even drop slightly, with each successive test (Test 1,  $M = .55, SE = .02$ ; Test 2,  $M = .55, SE = .02$ ; Test 3,  $M = .55, SE = .02$ ; Test 4,  $M = .54, SE = .02$ ), whereas recall performance for the participants in the uniform condition appeared to increase with each successive test (Test 1,  $M =$

**Table 1**  
**Mean Proportion of Critical Items Recalled During the Learning and Final Test Phases of Experiment 1 As a Function of Initial Study Time and Retrieval Practice Schedule**

| Condition                         | Initial Learning-Phase Tests |     |        |     |        |     |        |     | Final Delayed Tests |     |             |     |
|-----------------------------------|------------------------------|-----|--------|-----|--------|-----|--------|-----|---------------------|-----|-------------|-----|
|                                   | Test 1                       |     | Test 2 |     | Test 3 |     | Test 4 |     | Free Recall         |     | Cued Recall |     |
|                                   | M                            | SE  | M      | SE  | M      | SE  | M      | SE  | M                   | SE  | M           | SE  |
| Study Time: 5 min (Experiment 1A) |                              |     |        |     |        |     |        |     |                     |     |             |     |
| Expanding                         | .66                          | .03 | .67    | .03 | .67    | .03 | .67    | .03 | .45                 | .03 | .64         | .04 |
| Uniform                           | .63                          | .03 | .64    | .03 | .66    | .03 | .66    | .03 | .49                 | .03 | .68         | .04 |
| Study Time: 1 min (Experiment 1B) |                              |     |        |     |        |     |        |     |                     |     |             |     |
| Expanding                         | .44                          | .04 | .42    | .04 | .42    | .04 | .41    | .04 | .32                 | .04 | .48         | .04 |
| Uniform                           | .40                          | .04 | .38    | .04 | .40    | .04 | .43    | .04 | .33                 | .04 | .48         | .04 |

.52,  $SE = .02$ ; Test 2,  $M = .51$ ,  $SE = .02$ ; Test 3,  $M = .53$ ,  $SE = .02$ ; Test 4,  $M = .54$ ,  $SE = .02$ ). In fact, from Test 1 to Test 4, performance decreased an average of 1.1% in the expanding condition, whereas performance increased an average of 2.7% in the uniform condition. An independent samples  $t$  test confirmed that this difference was significant [ $t(86) = 1.99$ ,  $p = .05$ ].

**Recall after 1 week.** Performance on the final free recall test administered after the 1-week delay is shown in Table 1. Once again, although performance was lower overall for the participants given 1 min than for those given 5 min to study the passage, study time did not interact with any other variable in the experiments. As such, the analyses reported below collapse across participants in Experiments 1A and 1B.

As was true on the initial four tests, final free recall performance did not differ significantly as a function of retrieval-practice schedule [ $t(86) = 0.68$ ,  $p > .05$ ]. If anything, consistent with Karpicke and Roediger (2007), performance was better in the uniform condition ( $M = .42$ ,  $SE = .03$ ) than it was in the expanding condition ( $M = .39$ ,  $SE = .02$ ). A second  $t$  test was conducted to analyze performance on the cued recall task. Once again, however, performance in the uniform condition ( $M = .60$ ,  $SE = .03$ ) exceeded performance in the expanding condition, but not significantly ( $M = .57$ ,  $SE = .03$ ) [ $t(86) = 0.46$ ,  $p > .05$ ].

**Errors.** To explore errors made during free recall, we measured the extent to which the participants wrote information that contradicted each of the 15 critical facts. A written fact was only deemed an error if it directly contradicted one of the critical facts from the passage. During the initial four tests, the participants in the expanding condition contradicted essentially the same number of critical facts ( $M = 3.2\%$ ,  $SE = 0.6\%$ ) as did the participants in the uniform condition ( $M = 3.9\%$ ,  $SE = 0.6\%$ ), and although the participants tended to contradict slightly more facts after the 1-week delay in the expanding condition ( $M = 6.4\%$ ,  $SE = 1.2\%$ ) than in the uniform condition ( $M = 5.6\%$ ,  $SE = 0.9\%$ ), this difference was not statistically significant either [ $t(86) = 0.52$ ,  $p > .05$ ].

## EXPERIMENT 2

The results of Experiments 1A and 1B failed to demonstrate a benefit of expanding retrieval practice over uniform retrieval practice, and what (nonsignificant) differences there were at the time of final recall favored uniform retrieval practice. A problem, however, is that the to-be-learned information was not any more susceptible to forgetting between study and the initial test in the uniform condition than it was in the expanding condition. Performance on the initial test was essentially the same after 7 min of interpolated activity (uniform condition) as it was immediately (expanding condition). The participants in the uniform condition, therefore, may have benefited from the advantages of spaced testing without suffering from the disadvantages of increased test failures.

In Experiment 2, we introduced a powerful source of forgetting between initial study and the first test: interfer-

ence. Although the relationship between forgetting and the passage of time may seem strong, we know from a long history of research (e.g., McGeoch, 1932) that forgetting is related more strongly to the nature of interpolated activities or changes in contextual cues than to the actual passage of time (for a review, see Bjork, 2003). When multiple items in memory are associated to the same retrieval cue, recall of a particular item can suffer competition from other items, thereby producing forgetting (see, e.g., Anderson & Neely, 1996; Postman, 1971). The other items may have been learned before or after the target item (proactive and retroactive interference, respectively), and the degree of interference is a function of intertask similarity across learning episodes.

During the intertrial intervals of Experiments 1A and 1B, the participants read a passage about American government and politics. Although this was a verbal task, the material being learned was different in kind from the information learned about Antarctica, meaning that reading the passage may have failed to interfere with—and cause the forgetting of—information about Antarctica. In Experiment 2, we altered the interpolated task to make it more interfering. Rather than having participants read a passage about American government and politics, we asked them to learn about 10 new regions of the world in the same way that they had learned about Antarctica. By increasing intertask similarity and thereby increasing the degree to which the interpolated task interfered with initial learning, we expected the information about Antarctica to become much more vulnerable to forgetting, which we also expected would favor the expanding condition, as in Landauer and Bjork's (1978) original study.

## Method

**Participants.** A total of 30 undergraduate students (6 male and 24 female) from the University of California, Los Angeles, participated in the experiment for credit in an introductory psychology course. The students were on average 19.0 years old.

**Materials and Procedure.** The materials and procedure were nearly identical to those employed in Experiments 1A and 1B. One difference was that all of the participants were given 1 min to study the Antarctica passage. A second and more important difference concerned the particular task in which the participants engaged during the intertrial intervals. Rather than to read a passage about American government and politics, the participants were asked to read about 10 additional regions of the world (i.e., Siberia, Norway, Australia, Africa, Greenland, Hawaii, Ukraine, Canada, China, and Madagascar). Each of the distractor passages was created so as to have the same formatting and type of information as the Antarctica passage. Each passage, for example, was presented on a single  $8 \times 11$  in. piece of paper and separated into four paragraphs describing that region's geography, climate, people, and location. The 10 distractor passages were stapled together in a packet and given to the participants for study during the intertrial intervals.

## Results

**Learning-phase recall.** The proportion of critical facts recalled correctly on the initial four tests is shown in Table 2 as a function of whether the participants were tested with an expanding test schedule or a uniform test schedule. Recall performance for the initial four tests was subjected to a 4 (Test 1 vs. Test 2 vs. Test 3 vs. Test 4)  $\times$  2

**Table 2**  
**Mean Proportion of Critical Items Recalled During the Learning and Final Test Phases**  
**of Experiment 2 As a Function of Retrieval Practice Schedule**

| Condition | Initial Learning-Phase Tests |           |          |           |          |           |          |           | Final Delayed Tests |           |             |           |
|-----------|------------------------------|-----------|----------|-----------|----------|-----------|----------|-----------|---------------------|-----------|-------------|-----------|
|           | Test 1                       |           | Test 2   |           | Test 3   |           | Test 4   |           | Free Recall         |           | Cued Recall |           |
|           | <i>M</i>                     | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i>            | <i>SE</i> | <i>M</i>    | <i>SE</i> |
| Expanding | .51                          | .04       | .48      | .04       | .50      | .04       | .50      | .04       | .43                 | .04       | .54         | .05       |
| Uniform   | .25                          | .04       | .28      | .04       | .27      | .04       | .27      | .04       | .19                 | .03       | .33         | .03       |

(expanding vs. uniform) mixed design ANOVA, with test schedule serving as a between-subjects variable. Unlike in Experiments 1A and 1B, the participants in the expanding condition ( $M = .50$ ,  $SE = .04$ ) performed significantly better than did the participants in the uniform condition ( $M = .27$ ,  $SE = .04$ ) [ $F(1,28) = 17.23$ ,  $MS_e = 17.23$ ,  $p < .001$ ]. We succeeded, therefore, in creating conditions in which more forgetting occurred prior to the initial test in the uniform condition than prior to the initial test in the expanding condition. Whereas the participants in the uniform condition (who were tested after 7 min) recalled 25% of the critical facts on the initial test, the participants in the expanding condition (who were tested immediately) recalled 51% of the critical facts on the initial test. As in Experiments 1A and 1B, performance averaged over the schedule conditions was about the same across the four initial tests. As was the case in Experiments 1A and 1B, however, performance appeared to stay the same or decrease slightly from Test 1 to Test 4 in the expanding condition, whereas there was a small, nonsignificant increase from Test 1 to Test 4 in the uniform condition [ $t(28) = 1.44$ ,  $p = .16$ ].

**Recall after 1 week.** An independent samples  $t$  test was conducted on free recall performance after the 1-week delay. As is shown in Table 2, the participants in the expanding condition ( $M = .43$ ,  $SE = .04$ ) not only outperformed the participants in the uniform condition ( $M = .19$ ,  $SE = .03$ ) [ $t(28) = 4.69$ ,  $p < .001$ ], but they did so by more than a two to one margin. In a second independent samples  $t$  test, we found that this effect was observed on the subsequent test of cued recall as well ( $M = .54$  vs.  $M = .33$ ) [ $t(28) = 3.57$ ,  $p = .001$ ]. It appears, therefore, that under conditions where to-be-learned information is vulnerable to forgetting, expanding retrieval practice can lead to far superior long-term retention than can uniform retrieval practice.

**Errors.** To explore errors during recall, we measured the extent to which the participants recalled information that contradicted each of the 15 critical facts. During the initial four tests, the participants in the expanding condition contradicted 1.9% ( $SE = 1.2\%$ ) of the critical facts, whereas the participants in the uniform condition contradicted 5.9% ( $SE = 1.2\%$ ) of the critical facts [ $F(1,28) = 5.56$ ,  $MS_e = 0.05$ ,  $p < .05$ ]. This difference persisted across the final retention interval. On the final test, the participants in the expanding condition ( $M = 2.7\%$ ,  $SE = 1.3\%$ ) contradicted significantly fewer critical facts than did the participants in the uniform condition ( $M = 7.6\%$ ,  $SE = 1.7\%$ ) [ $t(28) = 2.30$ ,  $p < .05$ ].

### EXPERIMENT 3

The primary purpose of Experiment 3 was to replicate the interaction observed between Experiments 1A and 1B and Experiment 2. More specifically, the purpose was to show that expanding retrieval practice can enhance long-term retention relative to uniform spaced practice, but only when participants engage in an interpolated task that interferes with memory for the to-be-learned information. Once again, participants were first instructed to study a passage about Antarctica. Unlike in Experiments 1A, 1B, and 2, however, retrieval practice schedule was manipulated within subjects. All of the participants received expanding retrieval practice for one set of information from the passage and uniform retrieval practice for another set of information from the passage. The tests were altered in order to accommodate the within-subjects manipulation of retrieval practice schedule. Whereas the participants were asked to free recall facts about the passage in Experiments 1A, 1B and 2, the participants in Experiment 3 were given cued recall fill-in-the-blank questions to answer. Some questions were repeatedly asked using an expanding schedule, whereas other questions were repeatedly asked using a uniform schedule.

#### Method

**Participants.** A total of 34 undergraduate students (26 female, 8 male) from the University of California, Los Angeles, participated for course credit in an introductory psychology course. The students were on average 20.2 years old.

**Materials.** The materials were drawn from the same materials used in Experiments 1A, 1B, and 2. The participants first studied the Antarctica passage. Unlike in the prior experiments, however, the passage was presented as one long paragraph for the participants to study. Fill-in-the-blank questions were created for 12 of the 15 critical facts used in the first two experiments. The fill-in-the-blank questions consisted of sentences with one or two words missing (e.g., *The coldest temperature ever recorded at the South Pole is \_\_\_\_\_*). The 12 fill-in-the-blank questions were then divided into two sets of six questions. Each participant was repeatedly tested on one set using an expanding schedule and the other set using a uniform schedule, with the particular assignment counterbalanced across participants. Finally, the distractor materials were the same as those used in the prior experiments. Half of the participants studied 10 other regions of the world (interfering condition), whereas the other half of the participants read about the American constitution (noninterfering condition).

**Procedure.** The participants were given 1 min to study the Antarctica passage. Once the initial study phase was complete, the participants answered a series of fill-in-the-blank questions interleaved with the study of interpolated material. Each fill-in-the-blank question was presented on the computer screen for 8 sec, and the participants were asked to say the critical missing word(s) out loud for

the experimenter to record. There were four blocks of questions; the first block took place immediately after studying the passage, and each subsequent block took place 7 min after the completion of the previous block. The questions in the expanding condition were tested immediately following study and then twice more, after 7- and 14-min periods of studying the interpolated material (Blocks 1, 2, and 4, respectively). Questions in the uniform condition were tested after an initial 7-min period of studying the interpolated material, after an additional 7-min interval, and after a final 7-min interval (Blocks 2, 3, and 4, respectively). As such, Block 1 consisted of questions only in the expanding condition, Block 3 consisted of questions only in the uniform condition, and Blocks 2 and 4 consisted of questions in both the expanding and the uniform conditions. The order of the questions in each block was determined randomly, and the total time of interpolated study was 21 min in both the expanding and the uniform conditions. After completing the final block of fill-in-the-blank questions, the participants were informed that the first phase of the experiment was complete and that they should return in 1 week to complete the experiment. When they returned, each participant was tested again with all 12 critical fill-in-the-blank questions, presented in a random interleaved order.

## Results

**Learning-phase recall.** The proportion of critical facts recalled correctly on the initial three tests were subjected to a 3 (Test 1 vs. Test 2 vs. Test 3)  $\times$  2 (expanding vs. uniform)  $\times$  2 (interfering vs. noninterfering) mixed design ANOVA, with interference serving as a between-subjects variable. Overall, items tested in the expanding condition ( $M = .61$ ,  $SE = .05$ ) were recalled significantly better than items tested in the uniform condition ( $M = .45$ ,  $SE = .04$ ) [ $F(1,32) = 13.03$ ,  $MS_e = 1.24$ ,  $p < .001$ ]. More important, the benefit of expanding practice over uniform practice was significantly greater for the participants who studied interfering material between tests ( $M = .58$ ,  $SE = .07$ , vs.  $M = .33$ ,  $SE = .06$ ) than it was for the participants who studied noninterfering material between tests ( $M = .64$ ,  $SE = .06$ , vs.  $M = .57$ ,  $SE = .06$ ) [ $F(1,32) = 4.27$ ,  $MS_e = .41$ ,  $p < .05$ ]. Looking at the initial test alone, we see that performance in the expanding condition was identical in the interfering ( $M = .65$ ,  $SE = .07$ ) and noninterfering ( $M = .65$ ,  $SE = .06$ ) conditions, whereas performance on the initial test in the uniform condition varied substantially between the interfering ( $M = .34$ ,  $SE = .06$ ) and noninterfering ( $M = .55$ ,  $SE = .06$ ) conditions. Recall performance in each condition across all tests is shown in Table 3.

**Table 3**  
Mean Proportion of Fill-in-the-Blank Questions Answered Correctly During the Learning and Final Test Phases of Experiment 3 As a Function of Retrieval Practice Schedule and Type of Interpolated Material Studied

| Learning Condition                              | Test Trial |           |          |           |          |           |            |           |
|---|------------|-----------|----------|-----------|----------|-----------|------------|-----------|
|   | Test 1     |           | Test 2   |           | Test 3   |           | Final Test |           |
|   | <i>M</i>   | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i>   | <i>SE</i> |
| Interfering Material (Other World Regions)      |            |           |          |           |          |           |            |           |
| Expanding                                       | .65        | .07       | .54      | .07       | .55      | .07       | .48        | .07       |
| Uniform   | .34        | .07       | .30      | .06       | .36      | .06       | .29        | .07       |
| Noninterfering Material (American Constitution) |            |           |          |           |          |           |            |           |
| Expanding                                       | .65        | .06       | .63      | .07       | .63      | .07       | .48        | .07       |
| Uniform   | .55        | .06       | .58      | .06       | .59      | .06       | .54        | .06       |

**Recall after 1 week.** Performance on the final cued recall task was subjected to a 2 (expanding vs. uniform)  $\times$  2 (interfering vs. noninterfering) mixed design ANOVA, with interference serving as a between-subject variable. As is shown in Table 3, a significant interaction emerged, such that the participants in the interfering condition, who read about other regions of the world during the interpolated intervals, demonstrated a significantly larger benefit from expanding retrieval practice than did the participants in the noninterfering condition [ $F(1,32) = 6.81$ ,  $MS_e = 0.25$ ,  $p < .05$ ]. Whereas the participants in the interfering condition demonstrated a substantial advantage from expanding retrieval practice ( $M = .48$ ,  $SE = .07$ ) over uniform retrieval practice ( $M = .29$ ,  $SE = .07$ ) [ $t(15) = 3.75$ ,  $p < .01$ ], the participants in the noninterfering condition actually demonstrated a nonsignificant advantage from uniform retrieval practice ( $M = .54$ ,  $SE = .07$ ) over expanding retrieval practice ( $M = .48$ ,  $SE = .06$ ) [ $t(17) < 1$ ]. Expressed differently, these data suggest that studying interfering material between repeated tests only led to forgetting when those tests were scheduled with uniform interpolated intervals. When an expanding schedule was employed, the participants recalled exactly the same amount of information after a 1-week delay, regardless of the nature of the intervening material.

**Errors.** Next, we measured the extent to which the participants recalled information that contradicted each of the critical facts. During the initial three fill-in-the-blank tests, the participants in the interfering condition contradicted significantly more information ( $M = .20$ ,  $SE = .03$ ) than did the participants in the noninterfering condition ( $M = .13$ ,  $SE = .02$ ) [ $F(1,31) = 3.82$ ,  $p = .06$ ]. Although the interaction did not reach significance, this difference was greater in the uniform condition ( $M = .25$ ,  $SE = .04$ , vs.  $M = .13$ ,  $SE = .03$ ) than in the expanding condition ( $M = .15$ ,  $SE = .03$ , vs.  $M = .12$ ,  $SE = .03$ ) [ $F(1,31) = 2.60$ ,  $MS_e = 0.10$ ,  $p = .12$ ]. When error rates were measured on the final test, a quite interesting interaction emerged. Whereas the participants in the interfering condition made more errors following uniform practice ( $M = .27$ ,  $SE = .04$ ) than following expanding practice ( $M = .19$ ,  $SE = .05$ ), the participants in the noninterfering condition made more errors following expanding practice ( $M = .27$ ,  $SE = .04$ ) than following uniform practice ( $M = .18$ ,  $SE = .04$ ) [ $F(1,31) = 6.92$ ,  $MS_e = 0.12$ ,  $p < .05$ ].

## DISCUSSION

Testing and spacing have each been clearly shown to enhance the long-term retention of to-be-learned information and skills (see e.g., Bjork, 1994b, 1999; Roediger & Karpicke, 2006a). What is not clear is how the two manipulations can be most effectively combined. Even the one generalization that has seemed, on the basis of earlier data and arguments (see Bjork, 1988; Landauer & Bjork, 1978), to provide an important guideline for learners and instructors—namely, that a schedule of tests with increasingly spaced intertrial intervals represents the optimal method of combining spacing and testing—has been called into question (Balota et al., 2007). Recent ob-

servations that expanding retrieval practice leads to worse performance than uniform retrieval practice after a long retention interval (e.g., Karpicke & Roediger, 2007) suggest that the high levels of retrieval success induced by an expanding schedule may actually result in less efficient learning than does a uniform schedule of practice.

### When Expanding Retrieval Practice Is and Is Not Optimal

The results of the present experiments provide an explanation for the apparent discrepancies in the literature and provide a new, if somewhat more complicated, guideline for learners and instructors: Under conditions in which to-be-learned information is vulnerable to forgetting, expanding-interval retrieval practice can produce substantially better long-term recall than does uniform-interval practice. In Experiments 1A, 1B, and 2, the participants first read a passage about Antarctica and were then instructed to recall facts about Antarctica without feedback via either an expanding (0, 3, 7, and 18 min) or uniform (7, 7, 7, and 7 min) schedule of tests. In Experiments 1A and 1B, the interpolated activity between successive tests involved reading a passage about an unrelated/noninterfering topic (i.e., American politics and government). In Experiment 2, however, the interpolated activity between successive tests involved reading passages containing information about other regions of the world (e.g., Greenland, Africa, etc.), a task specifically designed to produce interference and therefore the forgetting of information from the Antarctica passage. Whereas there was not a significant difference in final recall performance after a 1-week delay in Experiments 1A and 1B, there was a very large benefit of expanding retrieval practice in Experiment 2. This interaction was replicated in Experiment 3 using fill-in-the-blank questions instead of free recall tests.

### Why Expanding Retrieval Practice Is Optimal, When It Is Optimal

The logic for why we observed this pattern of results is relatively straightforward. Expanding retrieval practice allows for spacing to occur between test trials while minimizing the costs associated with retrieval failures. By testing immediately and then systematically increasing delays between subsequent tests, expanding retrieval practice ensures that a maximum number of items will continue to be recallable and therefore continue to benefit from the powerful consequences of repeated testing. In the corresponding uniform schedule, the likelihood of retrieval failure on the first test is much higher, and any such failures will propagate to all subsequent tests. As in our Experiments 1A and 1B, however, when the participants in the uniform condition are able to recall approximately the same number of facts as the participants in the expanding condition, the benefit of expanding practice disappears.

There are reasons to expect expanding retrieval practice, under such conditions, to be inferior to uniform retrieval practice. Research has shown that increasing the delay between two learning trials can enhance the long-term retention of what is learned (for reviews of the spacing effect, see, e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006;

Dempster, 1996; Hintzman, 1974). A test given after a 7-min delay should therefore be more valuable than a test given immediately—especially if test performance in the two conditions is roughly equivalent. Because participants in the expanding condition do not benefit from the initial test to the same extent as participants in the uniform condition, it is perhaps not surprising that their performance on a final delayed test would suffer. Although spacing might increase in the latter trials of the expanding condition, the effectiveness of spacing during those latter trials may not make up for the ineffectiveness of the nonspaced initial test trial. The argument is similar to Landauer and Bjork's (1978) argument as to why they expected, and found, that uniform, not expanding, practice would be better for names repeatedly presented rather than tested:

In contrast, when the information is repeated, very long intervals are not as much better than moderate intervals as very short intervals are worse, so uniform spacing should be better for repetition-type practice. (p. 626)

When the intervening activity was designed to interfere with memory for the to-be-learned material, however, recall performance on the initial test differed markedly as a function of schedule. In Experiment 2, for example, the participants in the uniform condition recalled 25% of the critical facts, whereas the participants in the expanding condition recalled 51% of the critical facts. Thus, our attempt to create interference by manipulating the intervening activity was clearly successful. And because over twice as many facts were able to be recalled on the initial test in the expanding condition than in the uniform condition, over twice as many facts were able to benefit from the repeated tests that followed. That advantage then translated to over twice as many facts about Antarctica being recallable on the final test, 1 week later. The fact that the retrieval-practice benefits of an expanding schedule over a uniform schedule were fully maintained across a 1-week delay is itself both important and surprising. One might have anticipated that the retrieval-phase advantage of the expanding schedule would be reduced or eliminated after a 1-week retention interval. Nevertheless, this pattern of results was observed in Experiment 2 and replicated in Experiment 3.

One might wonder whether the probability of reexposure via successful recall is the principal benefit of expanding retrieval practice. The main reason for which we do not believe this to be the case is that prior research has demonstrated that the first interval can be too short as well as too long. Some of Karpicke and Roediger's (2007) results are such a demonstration, in our view, as are a number of Landauer and Bjork's (1978) initial findings. Landauer and Bjork found, for example, that when uniform schedules were contrasted, 0, 0, 0 and 1, 1, 1 schedules produced inferior long-term recall in contrast to 4, 4, 4 or 5, 5, 5 schedules, notwithstanding much higher Test 1 performance, and—perhaps more to the point—0, 3, 10 and 1, 4, 10 schedules produced much better final recall than did 0, 0, 0 and 1, 1, 1 schedules, respectively. We are also influenced by the research by Kornell and Bjork

(2008), who have addressed the issue more directly. In their research, they have tried to determine whether, after a first interval of a given length, the next interval should expand or not. If an optimal first interval is what matters, and increasingly spaced/difficult retrievals after that do not matter, expanding the intervals after the first test should be no better than keeping those intervals constant. Kornell and Bjork, however, found substantial benefits of expanding subsequent intervals relative to keeping them the same.

Finally, it is noteworthy that several researchers have failed to observe a significant advantage for expanding retrieval practice over uniform retrieval practice on the final test, despite showing a significant advantage during acquisition (e.g., Carpenter & DeLosh, 2005; Logan & Balota, 2008). For example, in the second experiment reported by Carpenter and DeLosh, performance during expanding retrieval practice was 20% better than performance during uniform retrieval practice, but was 4% worse on the final test. Similar results were observed in the non-interfering condition of Experiment 3 in the present study. Thus, it appears that expanding retrieval practice may not always be superior to uniform retrieval practice, even if it is effective at maintaining a higher level of performance during learning.

One possible explanation is that the benefits of expanding retrieval practice diminish as the final retention interval increases. That is, once retrieval practice ceases to continue, the advantage for the expanding condition may become progressively smaller and eventually even reverse. There are several reasons to expect this shift to occur. First, as was discussed above, the benefits of having a delayed initial test may become increasingly important as the retention interval itself is delayed. Furthermore, although expanding retrieval practice may be able to keep more items accessible during acquisition, whether those items are retained across a final retention interval may depend on a number of factors. For instance, if there are too few retrieval practice opportunities, sufficient learning may not occur, and the items whose accessibility was maintained during expanding retrieval practice may be lost. Of course, a virtue of expanding retrieval practice is that practice can continue indefinitely with increasingly sparse episodes of practice. Thus, provided retrieval practice continues, expanding retrieval practice has the potential to facilitate performance both in the short run and in the long run.

### **What About Other Ways to Induce Forgetting?**

Although increasing intertask interference may increase the effectiveness of expanding retrieval practice, any manipulation that makes the to-be-learned information more vulnerable to forgetting should have a similar effect. It is possible, for example, that simply increasing the overall amount of spacing between trials would be sufficient. Although a significant amount of forgetting did not occur after 7 min in the first experiment, it is possible that a substantial amount of forgetting would have occurred after 70 min. Thus, if the expanding condition

involved tests scheduled with 0-, 30-, 70-, and 180-min intertrial intervals, and the uniform condition involved tests with 70-, 70-, 70-, 70-min intertrial intervals, long-term performance might have been significantly better in the expanding condition than in the uniform condition, even if an interfering interpolated task was not introduced.

### **What If an Immediate Test Is Given in Both the Expanding and the Uniform Conditions?**

A fundamental virtue of expanding retrieval practice is that it keeps retrieval practice successful. Learners can introduce an increasingly spaced schedule of practice without losing access to the to-be-learned information. But what if the participants are given an immediate test in both the uniform and expanding retrieval practice conditions? In other words, consider an experiment in which both conditions involve an immediate initial test and then an expanding or a uniform schedule of tests that follow. Whether a subsequent schedule of expanding retrieval practice will be superior to a subsequent schedule of uniform retrieval practice will still depend on how vulnerable the to-be-learned information is to forgetting, only now it will depend on how vulnerable the information is to forgetting between the initial test and the second test. Of course, having had the opportunity to successfully retrieve the to-be-learned information on the first test would likely reduce that information's vulnerability to forgetting, but a long enough delay and/or sufficient interference could create the conditions for such forgetting to occur. Said differently, one can decide at any point how to schedule future retrieval practice trials. Whether immediately following learning, a first test, or an  $n$ th test, if the to-be-learned information is vulnerable to being forgotten, that information will benefit from the implementation of an expanding schedule of retrieval practice.

### **Error-Reduction Benefits of Expanding Retrieval Practice**

In addition to facilitating recall, expanding retrieval practice may also prevent the production and persistence of errors. To the extent that spacing induces forgetting, it also increases the likelihood of recalling incorrect information. And although research has shown that when learners are given feedback, the benefits of spacing can overwhelm the potentially harmful effect of generating incorrect information (e.g., Pashler, Zarow, & Triplett, 2003), when feedback is not given, the repeated retrieval of incorrect information may propagate those errors into the future (e.g., Toppino & Brochin, 1989). By testing immediately, and then systematically introducing spacing between repeated tests, it is possible and likely that the number of errors can be minimized.

Consistent with this expectation, the participants in the expanding condition, relative to those in the uniform condition, recalled significantly fewer erroneous items of information—that is, items that contradicted the critical facts in the passage. Thus, when interpolated with interfering material, expanding retrieval practice both facilitated the retention of correct information and diminished the

tendency for the participants to recall false information. In Experiment 2, whereas the participants in the uniform condition recalled merely 2.5 correct facts for each incorrect fact, the participants in the expanding condition recalled 16.2 correct facts for each incorrect fact. Although not as striking, this same pattern of results was observed in the interference condition of Experiment 3. When answering fill-in-the-blank questions on the final test, the participants generated only 1.1 correct answers for each incorrect answer on questions previously practiced with a uniform schedule of tests, whereas the participants generated 2.5 correct answers for each incorrect answer on questions previously practiced with an expanding schedule of tests. This is a stunning difference, especially given the importance in educational contexts of *both* preventing the acquisition of false information and facilitating the retention of true information.

## Conclusion

Until recently, expanding retrieval practice has widely been assumed to provide an optimal method of scheduling tests. Recent results, however, have called this assumption into question, showing that when memory is tested after a long retention interval, uniform retrieval practice leads to superior performance than expanding retrieval practice. The experiments reported here demonstrate that expanding retrieval practice can indeed—under some conditions—lead to superior long-term retention. Under conditions in which the to-be-learned material is vulnerable to being forgotten, scheduling tests with expanding intervals can promote both the successful recall of true information and prevent the unwanted recall of false information.

## AUTHOR NOTE

This research was supported by Collaborative Activity Grant 29192G from the James S. McDonnell Foundation. The results of Experiments 1 and 2 were reported at the meeting of the American Psychological Association, August 2007. We thank Elizabeth Ligon Bjork and the members of Cogfog for their valuable input and comments, as well as Joey Chen, Michael Friedman, and Jason Finley. Correspondence concerning this article should be addressed to B. C. Storm, Department of Psychology, University of Illinois at Chicago, 1007 W. Harrison St. (MC 285), Chicago, IL 60607-7137 (e-mail: bstorm@uic.edu).

## REFERENCES

- ANDERSON, M. C., & NEELY, J. H. (1996). Interference and inhibition in memory retrieval. In E. C. Carterette & M. P. Friedman (Series Eds.) and E. L. Bjork & R. A. Bjork (Vol. Eds.), *Handbook of perception and cognition: Vol. 10. Memory* (2nd ed., pp. 237-313). San Diego: Academic Press.
- BALOTA, H. P., DUCHEK, J. M., & LOGAN, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 83-106). New York: Psychological Press.
- BALOTA, D. A., DUCHEK, J. M., SERGENT-MARSHALL, S. D., & ROEDIGER, H. L., III (2006). Does expanded retrieval produce benefits over equal interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology & Aging*, *21*, 19-31.
- BJORK, R. A. (1975). Retrieval as a memory modifier. 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: Memory in everyday life* (pp. 396-401). New York: Wiley.
- BJORK, R. A. (1994a). Institutional impediments to effective training. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 295-306). Washington, DC: National Academy Press.
- BJORK, R. A. (1994b). 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.
- BJORK, R. A. (2003). Interference and forgetting. In J. H. Byrne (Ed.), *Encyclopedia of learning and memory* (2nd ed., pp. 268-273). New York: Macmillan.
- CAMP, C. J. (2006). Spaced retrieval: A model for dissemination of a cognitive intervention for persons with dementia. In D. K. Attix & K. A. Welsh-Bohmer (Eds.), *Geriatric neuropsychology: Assessment and intervention* (pp. 275-292). New York: Guilford.
- CARPENTER, S. K., & DELOSH, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619-636.
- CARRIER, M., & PASHLER, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633-642.
- CEPEDA, N. J., PASHLER, H., VUL, E., WIXTED, J. T., & ROHRER, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380.
- CULL, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215-235.
- CULL, W. L., SHAUGHNESSY, J. J., & ZECHMEISTER, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence of applicability. *Journal of Experimental Psychology: Applied*, *2*, 365-378.
- DEMPSTER, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. C. Carterette & M. P. Friedman (Series Eds.) and E. L. Bjork & R. A. Bjork (Vol. Eds.), *Handbook of perception and cognition: Vol. 10. Memory* (2nd ed., pp. 317-344). San Diego: Academic Press.
- DEMPSTER, F. N., & PERKINS, P. G. (1993). Revitalizing classroom assessment: Using tests to promote learning. *Journal of Instructional Psychology*, *20*, 197-203.
- GLOVER, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392-399.
- HINTZMAN, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77-99). Potomac, MD: Erlbaum.
- 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.
- JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, *17*, 649-667.
- KARPICKE, J. D., & ROEDIGER, H. L., III (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *33*, 704-719.
- KORNELL, N., & BJORK, R. A. (2008, November). *Expanded retrieval practice in theory and practice*. Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.
- LANDAUER, T. K., & BJORK, R. A. (1978). Optimal 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.
- LOGAN, J. M., & BALOTA, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, & Cognition*, *15*, 257-280.
- MCDANIEL, M. A., & MASSON, M. E. J. (1985). Altering memory rep-

- representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 371-385.
- MCGEOCH, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, **39**, 352-370.
- MODIGLIANI, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 609-622.
- 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, & Cognition*, **29**, 1051-1057.
- POSTMAN, L. (1971). Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), *Woodsworth and Schlosberg's experimental psychology* (3rd ed., pp. 1019-1132). New York: Holt, Rinehart & Winston.
- REA, C. P., & MODIGLIANI, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, **4**, 11-18.
- 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.
- SCHACTER, D. L., RICH, S. A., & STAMPP, M. S. (1985). Remediation of memory disorders: Experimental evaluation of the spaced-retrieval technique. *Journal of Clinical & Experimental Neuropsychology*, **7**, 79-96.
- SPITZER, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, **30**, 641-656.
- TOPPINO, T. C., & BROCHIN, H. A. (1989). Learning from tests: The case of true-false examinations. *Journal of Educational Research*, **83**, 119-124.
- TULVING, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, **6**, 175-184.
- 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.
- WHITTEN, W. B., II, & BJORK, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 465-478.
- WILSON, B. A., BADDELEY, A. D., EVANS, J., & SHEIL, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation*, **4**, 307-326.

(Manuscript received March 24, 2009;  
revision accepted for publication August 7, 2009.)