The Journal of General Psychology, 2013, *140*(3), 224–241

The Effect of a Final Exam on Long-Term Retention

ARNOLD LEWIS GLASS
MARGARET INGATE
NEHA SINHA
Rutgers University

ABSTRACT. Testing on a final exam in a college course improved long-term retention over material that had not been tested on the final. Students from an upper level psychology course took a long-term retention test, four to five months after the end of the course. For half of the items, a related question had been on the final. For the remaining half, a related question had appeared on an earlier exam, but not the final. On the long-term retention test, percent correct was 79% when a related question had appeared on the final and 67% when a related question had not appeared on the final. These results have both theoretical and practical implications.

Keywords: distributed study, generalization, testing effect

THE RELATIONSHIP BETWEEN VOLUNTARY ACTION and learning is of theoretical importance to the understanding of learning and of practical importance to pedagogy. Specifically, at the very beginning of experimental psychology, Ebbinghaus (1985) found that retrieving study material caused savings during relearning of that material and the basic finding that spaced (or distributed) retrieval practice results in better retention than massed practice has been replicated many times since (e.g., Cook, 1944; Ericksen, 1941; Underwood, 1961). It predicts that making a study item the target of a voluntary action will improve understanding and increase retention over a condition in which the study item is seen or heard with no specific response required.

The specific retrieval task investigated here was answering a question about the study material. The finding that answering a question on study material increases its retention compared with repeated reading or listening is called the testing effect. There is a long, but sporadic, history of research confirming the testing effect (Gates 1917; Rothkopf, 1966; Spitzer, 1939; Zangwill, 1939). Hamaker (1986)

Address correspondence to Prof. Arnold Lewis Glass, Rutgers University, Psychology, 152 Frelinghuysen Rd., Piscataway, NJ 08844, USA; aglass@rutgers.edu (e-mail).

reviewed all well-controlled experiments that investigated the effects of questions on prose learning performed up to that time and found that question-answering improved retention.

Furthermore, spacing the questions for a single fact statement over intervals during which a student answers other questions or performs completely different tasks increases the effect of spaced (or distributed) questioning on learning and retention compared with massed questioning (Roediger & Karpicke, 2006a, 2000b; Rohrer & Taylor, 2006; Pashler, Zarow, & Triplett, 2003).

Dual-System Theory

A variety of neural mechanisms have been discovered that result in better retention from massed than distributed experience, beginning with the role of protein synthesis observed in synaptic transmission in Aplysia (Jin, Kandel, & Hawkins, 2011; Naqib, Farah, Pack, & Sossin, 2011; Sutton, Ide, Masters, & Carew, 2002). Among these, the effect of distributed questioning on long-term retention is best explained by the dual-system theory of mammalian memory. As reviewed by Yin and Knowlton (2006), mammalian memory appears to consist of two qualitatively different memory systems: the instrumental system, whose neural pathway includes the hippocampus, and the habit system, whose neural pathway includes the striatum. The instrumental memory system includes the representation of the environment by mental maps and controls the representation of novel targets. When an action directed towards a novel target is successful, a fragile representation within the instrumental system is encoded. The habit memory system includes the representation of a sequence of goal directed actions, e.g., the sequences of turns that must be made to navigate a familiar maze. When a novel target is repeated and recognized through a match with its instrumental representation, and the same action directed towards it is again successful, a robust representation within the habit system is encoded (Sadeh, Shohamy, Rubi Levy, Reggev, & Maril, 2011). An influential study by Packard and McGaugh (1996) found that one week after training, a rat's response in a T-maze depended on the hippocampus, but two weeks after training, the rat's response depended on the striatum. Consistent with the hypothesis of discrete short- and long-term memory systems operating in parallel, Izquierdo, Medina, Vianna, Izquierdo, and Barro (1999) found that the same chemical treatments selectively interfered with either short- or long-term memory depending on where they were applied in the hippocampus, anterolateral prefrontal cortex, or amygdala.

As the precise functions of the two systems have been revealed over recent decades, they have also been called the declarative and procedural systems (Squire, 1986), explicit and implicit memory (Fu, Bin, Dienes, Fu, & Gao, 2013) and the declarative and habit systems (Sadeh et al., 2011). What all the pairs of terms have in common is that they represent two different senses of the word "memory." Declarative and explicit refer to the awareness that a task has been performed

or a target perceived before. This is also called "knowing what" as in knowing what to do or what something is. Procedural and implicit refer to an improvement in performance as a result of practice. This improvement does not an imply that the task has been performed before. This is also called "knowing how." However, the terms explicit memory and implicit memory refer to observable behaviors without implying a causal mechanism for those behaviors. For example, they do not imply whether they are the result of one memory system or two. In contrast, Squire (1986) used the terms declarative and procedural to identify two distinct memory systems, but at that time the domains and boundaries of the memory systems were not understood. In contrast, the terms instrumental system and habit system specifically identify well-defined neural circuits and their functions (Yin & Knowlton, 2006).

The purpose of this two-system architecture is to preserve memories of both highly emotional events and routine events. When a person performs a voluntary action, an episode is encoded in memory consisting of the target of the action, the action itself, and the result of the action. The instrumental system appears to be sensitive to the valence and intensity of the result of the action, i.e., the reward. Consequently, an action whose result includes a strong emotional response produces a robust episode in a single trial that may be retained indefinitely. Conversely, an action directed towards a novel but mundane target results in a fragile episode. The habit system appears to be sensitive to the contingency between the action to the target and its result. Since only a weak contingent relationship can exist after only one trial, initially both the instrumental and habit representations of the episode are fragile (Yin & Knowlton, 2006). Hence, the general prediction derived from the two-system description of mammalian memory is that a single response to a novel mundane target is not sufficient to initiate the encoding of a robust representation of it (Viskontas, Carr, Engel, & Knowlton, 2009).

Consequently, a newly constructed episode that describes a novel but mundane experience is fragile, so that after a retention interval of several days it no longer exists. A response to a second occurrence of the target must be made during a critical time period that is long enough to be encoded as part of a different episode but not so long that its initial representation is no longer available in order to construct a robust representation. Only when the repetition of the target falls within this time window is a robust representation of it constructed (Rovee-Collier, Evancio, & Earley, 1995). When the repetition of a target is delayed until it is at the end of the time window for increasing the robustness of the representation of its first occurrence, the repetition is said to be distributed or spaced. When repetition occurs immediately after the first occurrence, (and hence should not increase retention) it is said to be massed.

The effect of distributed study in humans is observed in infants as young as three months of age (Rovee-Collier, 1993). According to the dual-system hypothesis, after a representation has been established in memory by an initial study episode, subsequently, it is only necessary to provide a reminder specific to that

study episode in order to access that representation and increase its robustness within the habit system. For example, if a 6-month-old infant sees one puppet pull a glove off another puppet, the infant will remember the event and be able to imitate the action one day but not two days later. However, if the infant is given the opportunity to imitate the event one day later, then the infant will remember the event and be able to imitate the action ten days later. Furthermore, if instead the infant is merely reminded of the event one day later by being briefly shown the puppet, then the infant will still remember the event and be able to imitate the action ten days later (Barr, Rovee-Collier, & Campanella, 2005). Hildreth and Rovee-Collier (1999) found that a brief reminder of as little as 15 seconds is as effective at increasing retention as an entire repeated episode of many minutes. If a single question is considered to be a reminder then asking a single question should similarly extend retention of its answer. Hence, distributed questioning should extend retention of the answers to those questions.

Furthermore, over the past 15 years, many studies had shown that spaced retrieval was effective at ameliorating the memory impairment associated with dementia by increasing retention of useful daily cognitive skills. Among them, Camp, Foss, O'Hanlon, and Stevens, (1996) found that spaced practice helped in maintenance of cognitive skills like calendar checking in order to keep track of one's schedule. Seifert and Baker found that spaced retrieval maintained collage-making skills (Seifert & Baker, 2002) and picture recognition skills (Seifert & Baker, 2009). Bourgeois, Camp, Rose, White, Malone, Carr, and Rovine (2003) found that 23 of 35 demented individuals were able to use spaced retrieval to learn daily retrieval skills and five of them retained their skills four months after training. Similarly, Sumowski, Wood, Chiaravalloti, Wylie, Lengenfelder, and Deluca (2010) used retrieval practice as a strategy for helping memory performance among persons with traumatic brain injury.

To summarize, there is considerable experimental evidence that distributed practice improves long-term retention in animals, distributed reminding improves long-term retention in infants, distributed retrieval improves long-term retention in demented and brain damaged individuals, and distributed testing improves short-term retention in children and adults. However, there are limited data about the effects of distributed testing on long-term retention for students learning academic materials. The primary reason for this is that, until recently, doing the experimental studies necessary to acquire this knowledge would have been prohibitively expensive. Hence, virtually all the research on distributed question-answering by college and younger students were laboratory studies that consisted of a single session or two sessions no more than a week apart. For example, Bloom and Shuell (1981) found that distributed questioning produced better retention than massed questioning for foreign vocabulary on a single quiz four days later.

More recently, an effect of distributed testing on long-term retention was found in classroom settings with students of various ages (Carpenter, Pashler, Cepeda, & Nicholas, 2009; McDaniel, Anderson, Derbish, & Morrisette, 2007;

Solity, 2000). However, these were limited tests that were not integrated into the curriculum that the students were studying. When Cepeda, Pashler, Vul, Wixted, and Rohrer (2006) examined more than 400 studies involving verbal recall, only about a dozen of these looked at retention intervals as long as one day, with just a handful examining retention intervals longer than one week.

As Cepeda et al. (2006) pointed out, there were only two pioneering studies involving long retention intervals. Bahrick, Bahrick, Bahrick, and Bahrick, (1993), and Bahrick and Phelps (1987) studied the acquisition and retention of foreign vocabulary over several years. However, in these studies at each session the vocabulary were relearned to criterion. So these studies investigated savings in re-learning from old knowledge rather than directly measuring the effect of distributed questioning on long-term retention.

Finally, to begin the study of the effect of distributed questioning on long-term retention, Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) made use of online technology to perform an innovative experimental study. Its purpose was to determine the optimal relationship between the (study) interval between the first and second times a question was asked and the (retention) interval between the second and third time the question was asked. A total of 1,354 individuals were taught a set of 32 obscure facts online. Each of 32 questions was repeated, with the correct answer as feedback, until answered correctly. After a study interval of up to 3.5 months each question was presented in an online session two times with the answer presented as feedback. After a retention interval of up to 1 year, each question was again presented in an online session. When measured as a proportion of retention interval, the study interval producing the greatest retention declined from about 20 to 40% of a one-week retention interval (2–3 days) to about 5–10% of a one-year retention interval (18–36 days). This important study demonstrated that distributed questioning could cause long-term retention. However, being a seminal study, it raised more questions than it answered. One inevitable limitation of a single study is that its results are limited to the kinds of study materials employed. Cepeda et al. (2008) used obscure facts, which were unlikely to be subject to retroactive interference from related knowledge because they were unrelated to anything else the participants knew.

Purpose of Experiment: Effect of a Final Exam on Post-Course Retention

To further study the effect of distributed questioning on long-term retention, Glass, Brill, and Ingate (2008) introduced the course-embedded experimental paradigm. Four recent experimental studies made use of the methodology to evaluate the effect of distributed questioning on performance in an actual academic course. Two of these courses were middle school courses (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011) and two of these courses were college courses (Glass, 2009; Glass et al., 2008).

The four studies made use of variants of the same experimental procedure, in which a multiple-choice exam question or a related question was asked up to three times before the exam. Throughout the course, pre-lesson questions were presented at the beginning of a class presenting the lesson containing the fact statement tested by the question. Post-lesson questions were presented shortly after the fact statements they tested at the end of the lesson. Review questions were presented two or more days after the class containing the lesson. The questions were presented as power point slides. Students used personal response devices (clickers) to answer them. Immediately after the class responded, a green checkmark indicated the correct answer. Furthermore, each exam question was presented on a monthly unit exam and again on the end-of-semester final exam. Most important, all four studies employed counter-balanced within-student, within-question experimental designs, embedded within multi-section courses, in which the experimental factor was the effect of the number of related questions prior to the exam on performance on the exam question. The results of all four studies were that presenting the identical or a related question at least twice before an exam increased the probability of answering the exam question correctly on both the unit exam and final exam. These results were consistent with those of Cepeda et al. (2008) and extended results for obscure, unrelated facts to the domain of useful academic knowledge for a retention interval of up to three months (the interval between the first unit exam and the end-of-the-semester final exam).

McDaniel et al. (2011) and Roediger et al. (2011) also included an end-of-the-year final exam. For some questions, this involved a retention interval of up to eight months. Over several experiments, McDaniel et al. (2011) consistently found that distributed questioning improved performance on the end-of-the-year exam but Roediger et al. (2011) did not. One possible explanation for the inconsistent results over the two studies is that students studied for the end-of-the-year exam and this influenced their performance. However, such study might have influenced exam performance in either of two opposite ways. On the one hand, if there was equal study of both material presented through distributed questioning and material not so presented just before the end-of-the-year exam, it might have obliterated the effect of distributed questioning in the Roediger et al. study. On the other hand, if there was more study of the material presented through distributed questioning, it might have caused the effect of distributed questioning in the McDaniel et al. study.

As mentioned above, Cepeda et al. (2008) found an effect of distributed questioning at a retention interval of a year using an experimental paradigm that discouraged study during the retention interval. This finding suggests that distributed questioning also played a role in the results of McDaniel et al. (2011) and Roediger et al. (2011). However, the context of Cepeda et al.'s result was quite different: participants learned novel unrelated facts in a novel experimental context. Hence, the result of Cepeda et al. cannot with confidence be generalized to the academic courses studied by McDaniel et al. and Roediger et al. and used to

determine the effect of an exam on long-term retention when there is no studying during the retention interval. This was the purpose of the experiment reported here.

In this study, students were told that the final exam would be cumulative and would cover only material covered by the three unit exams. In fact, twenty-one pairs of questions on the same topic and of the same difficulty were selected from the three unit exams and one question from each pair appeared on the final exam. Presumably, the students studied the answers to both the questions that did and did not appear on the final equally. Furthermore, there was no reason for the students to re-study the course material over the four-month interval after the end of the course.

Four months after the final exam, an email was sent to all the students who had participated in the course asking them to take a post-final retention test online. For both members of each of the 21 question-pairs, hence a total of 42 questions that each appeared on one of the three unit exams (E-questions), a new question was constructed whose answer was implied by the same fact statement that implied the answer to the final exam question (R-questions). The 42 R-questions comprised the retention test. Each E-R question pair whose answers were implied by the same fact statement will be called related questions. The retention test contained R-questions related to the E-questions rather than merely repeating the E-questions because the long-term retention of the fact statements tested by the E-questions, rather than the exact answers to the questions, is of more interest both theoretically and practically.

Hence, the retention test contained 21 pairs of R-questions such that one member of each pair was related to an E-question that appeared on the final exam and the other member of the pair was related to a similar E-question that did not appear on the final exam. Hence, the only difference in the previous study history for the members of each R-question pair was that for only one a related E-question had appeared on the final exam. So, better performance on the post-final retention test for the R-questions related to E-questions that appeared on the final exam could be attributed to the appearance of the E-questions on the final exam rather than any additional study for those questions. So, a positive result would demonstrate an effect of distributed questioning at a retention interval of 4 months.

Implications for Pedagogy

There are at least three possible reasons for giving a final exam beyond the obvious one of providing a summative measure of student performance. The first reason is that administering a final gives the students both an opportunity and an incentive to further study the course content and increase their knowledge. Students do make use of this opportunity to increase their knowledge. For example, Glass

(2009) found that that average percent correct increased from 74% to 86% when questions on unit exams were repeated on the final.

The second reason for giving a final is that distributed study increases post-course long-term retention (Bahrick et al., 1993; Bahrick & Phelps, 1987; Cepeda et al., 2008). Consequently, giving the students both an opportunity and an incentive to restudy the course content increases its long-term retention beyond the end of the course. Since long-term retention is virtually always the goal of academic instruction, the effect on long-term retention by itself justifies administering final exams.

The third reason for giving a final is that answering a question on the final itself, apart from any prior studying, increases long-term retention of the answer to the question (Cepeda et al. 2008). Additional recent studies have confirmed and extended the effect of distributed questioning on retention. Questioning was found to have a positive effect on retention even when feedback as to the correct answer was not given (Karpicke & Roediger, 2007).

Assessing the effect of the final, apart from prior studying, is important because it useful to know whether a final exam must be exhaustive to be maximally effective. If the effect of the final is only to encourage studying then there should be no difference in long-term retention between facts that were actually queried on the final and similar facts that were not. In that case, to increase the retention of all course content it would only be necessary to include some of the material on the final. The final would not have to be exhaustive. However, if actually being queried on the final increases long-term retention for a fact or concept, then in order to maximize the retention of course content, final exams should be comprehensive.

Method

Participants

This experiment made use of the students, study, and test materials of a psychology lecture course on memory completed by 371 students.

Organization of Course

The course was organized as follows. Twenty-seven lecture periods were partitioned into three blocks of nine periods each. Each block consisted of eight lecture classes followed by a unit exam containing 30 multiple-choice questions on the content of those eight lectures. So the unit exams were not cumulative. A week after the third unit exam there was a final exam during a three-hour final exam period.

The students were advised on the syllabus and in class by the instructor that the final exam would only review material covered on the three prior exams, but

they did not know that 21 questions from the prior exams would be repeated on the final exam.

During the three-hour final exam period, a 45-question final was given consisting of 15 questions from each 30-question unit exam. Twenty-four of the questions, 8 from each unit exam, were unrelated to the questions on the retention exam. Twenty-one of the questions, 7 from each unit exam, were each related to a question on the retention exam and matched with a similar question on the unit exam that did not appear on the final exam.

Subsequently, the retention exam four months after the final exam was composed of the 42 R-questions related to the 21 pairs of E-questions such that one member of each pair appeared on the final exam.

Materials

Twenty-one pairs of exam questions (E-questions) were selected, seven pairs from each unit exam, so that the two questions in the pair were on the same topic and had been answered correctly by almost exactly the same number of students. One member of each pair appeared among the 45 questions on the final exam. Consequently, there was not a significant difference in percent correct on the unit exams between the 21 E-questions that appeared on the final (79%) versus the corresponding 21 E-questions that did not appear on the final (81%).

For each one of the 42 E-questions, an R-question was constructed whose answer was implied by the same fact statement that implied the answer to the E-question. An example of an E-, R-question pair is shown below. If the E-question involved the application of a principle to an example then an entirely different example was selected for the R-statement. Otherwise, if the E-question involved a factual statement, its initial statement and all five alternatives were paraphrased to construct the R-statement. The two examples in the Appendix give the range of relationships between related E- and R-questions. The correct answer is shown in bold:

E. In order to determine that an infant remembers the training mobile the infant must be tested:
when his foot is **not** connected to the mobile
in the training room
with the crib liner from the training session around the crib
all of the above
none of the above

R. In the mobile recognition paradigm the infant is connected to the mobile during:
baseline
training
test
all of the above
none of the above

The Long-Term Retention Test

Each of the 42 R-questions appeared on the online retention test. The final was administered on December 19, and the long-term retention test became available April 28 and was available until May 16 of the following year. So, between 4 and 5 months elapsed between the final and retention test. Participation was solicited through an email to students who had taken the final exam. They received no compensation for participating in the online retention test.

Results and Discussion

A total of 55 students participated in the long-term retention test. Forty-five percent of the students taking the retention exam were male, which was the same percentage of those taking the final. The mean score on the final for those students taking the online retention exam was 83%, which was not significantly different from the class average on the final of 86%.

To assess the effect of the final on retention from a unit exam to the retention test, a 2×2 analysis of variance was used in which the fixed independent factors were Exam (Unit Exam versus Retention Test) and occurrence On Final (Yes versus No). The dependent measure was percent correct on the unit exams (for the E-questions) and on the retention exam (for the R-questions). The hypothesis that occurring on the final increased long-term retention predicted an interaction between Exam and On Final: that there would be a smaller decline in percent correct from the E-questions on the unit exams to the related R-questions on the retention exam for those questions for which the E-question also appeared on the final than for those questions for which the E-question did not also appear on the final.

The results are shown in Figure 1 and includes bars indicating 95% confidence errors. As shown in the figure, for E-questions that appeared on the final, mean percent correct was 80% on the unit exams and 79% for the related R-questions on the retention exam. For E-questions that did not appear on the final, mean percent correct was 86% on the unit exams and 67% for the related R-questions on the retention exam. The figure shows a smaller decline in performance from the unit exams to the retention test when the exam question also appeared on the final than when it did not, which was the predicted interaction. Recall that the 21 pairs of questions selected were such that the 375 students in the class answered 79% of the E-questions correctly that appeared on the final exam and 81% of the E-questions correctly that did not appear on the final exam. However, as shown in the figure, the subset of 55 students who took the retention test answered 80% of the E-questions correctly that appeared on the final exam but answered 86% of the E-questions correctly that did not appear on the final exam. Nevertheless, despite this better initial performance for the questions that did not appear on the final on the unit exams, forgetting was so great for them that performance was worse for them on the retention test, thus producing the interaction.

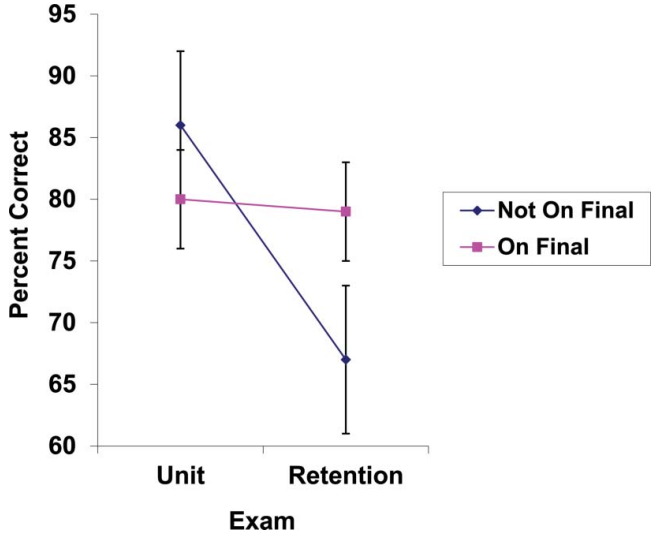

FIGURE 1. Mean percent correct (and 95% confidence intervals) showing the effect of a final question on subsequent performance on related questions four–five months later. (Color figure available online).

To assess the generalizability of the results to both different samples of students and different samples of questions, two analyses of variance were performed, as suggested by Clark (1973). In one analysis the random factor was subjects (students) and in the other analysis the random factor was items (questions). When subjects was the random factor, the effects of Exam, $F(1,54) = 17.3$, $d = .69$, and the Exam $\times$ On Final interaction, $F(1,54) = 20.5$, $d = 2.3$, were significant at the $p < .001$ and the effect of On Final, $F(1,54) = 8.2$, $d = 1.8$ was significant at the $p = .006$. Consistent with the significant interaction, a planned comparison between the change in percent correct from the unit to the retention exam for exam questions that did versus did not appear on the final was significant, $t(54) = 8.6$.

When questions was the random factor, the effects of Exam, $F(1,20) = 6.36$, and the Exam $\times$ On Final interaction, $F(1,20) = 4.44$, were significant at the $p < .05$ but the effect of On Final, $F(1,20) = .94$, was not significant, $p = .35$. The results of this analysis are shown in Figure 2. Min-F's were computed for which both subjects and items were random effects. By the min-F, Exam was significant at the $p < .05$ level, $F(1,36) = 4.67$, but the interaction was not, $F(1,29) = 3.62$.

This result demonstrates that when a question appears on a final exam, it sometimes increases long-term retention of its answer even when there has been considerable study of that answer prior to the final. As mentioned above, this has a

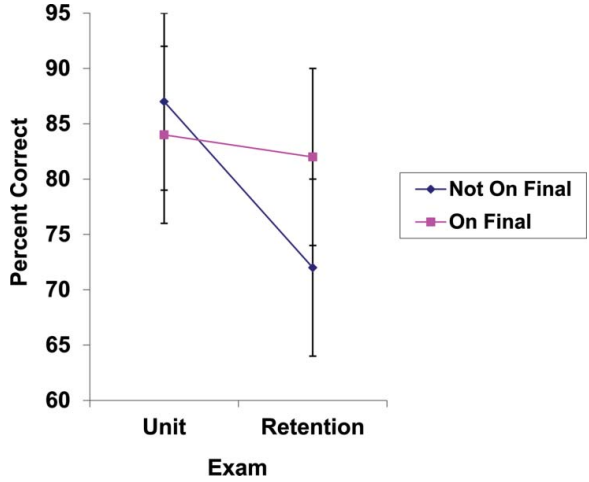

FIGURE 2. Mean percent correct (and 95% confidence intervals) showing the effect of a final question computed over items. (Color figure available online).

direct implication for pedagogy. In order to maximize retention of course content it appears that a final exam should be given, and, in some cases, it should be as comprehensive as practical.

Percent correct for the 21 pair members included on the final exam is shown in Table 1. As can be seen from the table, percent correct increases when the E-questions are repeated on the final. The improvement in performance is evidence that the students studied for the final exam. Percent correct declined four months later. So there was forgetting following the final exam. However, the final both raised the level of the performance and increased the robust of the memory, thus decreasing the amount of forgetting compared with questions not on the final exam, as shown in Figure 1.

TABLE 1. The Change in Mean Percent Correct (95% Confidence Intervals) for Questions Appearing on The Final Exam

Unit Exam	Final Exam	Retention Exam
80 (76, 85)	88 (86, 91)	79 (74, 84)

The result also has an implication for learning theory. Apparently a reminder, in the form of a question is sufficient to produce long-term retention. The effect of a question found here is consistent with the results of McDaniel et al. (2011). Therefore, the inconsistent results of Roediger et al. (2011) may be the result of studying for the end-of-the-year final sometimes obliterating the effect of a question's appearance on the final.

The statistical effect of the final was much weaker when analyzed over items than for subjects. To determine the reason for this, the individual item differences were examined. When an E-question appeared on the final, the decline in performance from the unit exam to the retention test was more than 10% for only one of the 21 items. That is, the occurrence of the E-question on the final served as an effective reminder that forestalled forgetting of the fact-statement implying its answer, so performance on the related R-statement was, at worst, only slightly less. In contrast, when an E-question did not appear on the final, the decline in performance from the unit exam to the retention test was more than 10% for 11 of the 21 items. Hence, the answers to 10 of the questions were learned well enough during the course to be retained four months after the course without being queried on the final. Because there was little forgetting for the answers to 10 questions that did not appear on the final, the statistical effect was weaker when items was treated as the random effect because it was the result of only 11 of the 21 items. However, this was not because the final failed to generally forestall forgetting, it was effective for 20 of 21 items. Rather, it was because studying an item in the absence of its appearance on the final by itself prevented forgetting for 10 of the 21 items. Hence, the weaker statistical result for items does not suggest that appearance on final is insufficient to produce long-term retention but rather that it is not always necessary.

The weaker effect over items cannot be attributed to an effect of the different intervals between each of the three unit exams and the final exam. When Unit was included as a factor in an analysis of variance, Exam, $F(1,54) = 25.5$, $p < .001$, and the Exam $\times$ On Final interaction, $F(1,54) = 26.2$, $p < .001$, On Final, $F(1,54) = 11.4$, $p = .002$ were again significant, as well as Unit, $F(2,108) = 33.4$, $p < .001$, and the Unit $\times$ On Final interaction, $F(2,108) = 8.5$, $p < .001$; however, the critical Unit $\times$ Exam $\times$ On Final interaction, $F(2,108) = .91$, $p = .405$, testing whether the effect of the final varied over the three unit exam-final exam intervals, was not significant. As can be seen by visual inspection of Figure 3, the same Exam $\times$ On Final interaction occurred for all three unit-final intervals.

A final exam that tests every single fact that it would be worthwhile for the student to retain is an inefficient way of creating long-term retention. One way of reducing the need for a comprehensive final is to create a comprehensive practice exam that students could take online in preparation for the final exam. However, such a comprehensive practice exam is no more efficient than a comprehensive final exam in terms of the students' total time spent answering questions. It merely displaces the time investment to the evening before the actual exam. A more

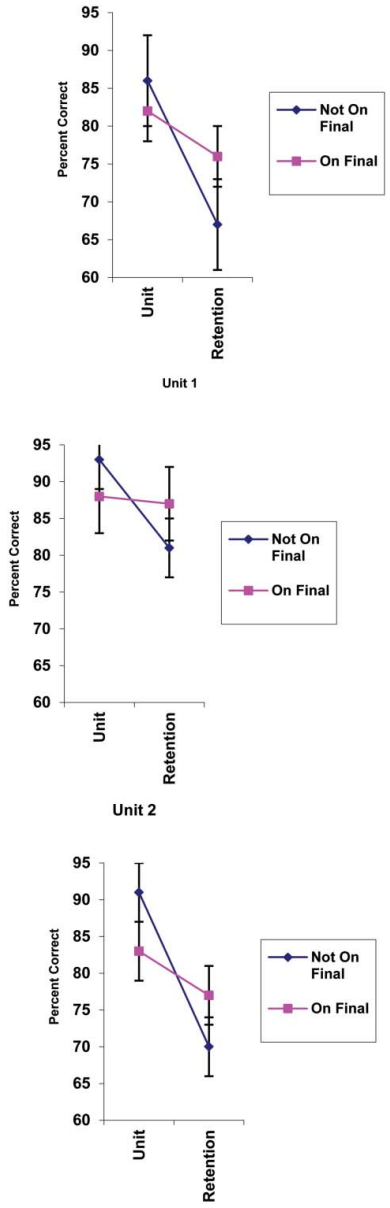

FIGURE 3. Mean percent correct (and 95% confidence intervals) showing the effect of a final question for questions on each unit exam. (Color figure available online).

promising approach might be to encourage the integration of knowledge so that a question on the final would be an effective reminder for more than one related question. Perhaps one reason that half the questions that did not appear on the final did not show a decline on the long-term retention test was that they were similar enough to questions that did appear on the final, so that students were reminded of their answers.

These results indicate that a final exam increases long-term retention for the material tested on the exam. This result was predicted by the dual system model of mammalian memory, which specifically predicts that distributed questioning increases long-term retention. There are not results from other studies on the effect of a final exam on long-term retention to compare this result with. However, several studies have found that distributed questioning increased long-term retention (e. g., Cepeda, et al., 2008; Glass, et al., 2008). These results indicate that it would worthwhile to further investigate the effect of a final on long-term retention.

AUTHOR NOTES

Arnold Lewis Glass is a professor in the cognitive program of the Psychology department at Rutgers University. **Margaret Ingate** is an instructor in the Psychology department at Rutgers University. **Neha Sinha** is a graduate student in the cognitive program of the Psychology department at Rutgers University.

REFERENCES

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316–321.

Bahrick, H.P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344–349.

Barr, R., Rovee-Collier, C., & Campanella, J. (2005). Retrieval protracts deferred imitation by 6-month-olds. *Infancy*, *7*, 263–283.

Bloom, K.C., & Shuell, T.J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*, 245–248.

Bourgeois, M. S. Camp C., Rose M., White, B., Malone, M., Carr, J., & Rovine, M. (2003). A comparison of training strategies to enhance use of external aids by a person with dementia. *Journal of Communication Disorders*, *36*, 361–378.

Camp, C. J., Foss, J. W., O'Hanlon, A. M., & Stevens, A. B. (1996). Memory interventions for persons with dementia. *Applied Cognitive Psychology*, *10*, 193–210.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology 23*, 760–771.

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.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning. *Psychological Science*, *19*, 1095–1102.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Cook, T.W. (1944). Factors in massed and distributed practice. *Journal of Experimental Psychology*, *34*, 325–334.

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. New York, NY: Dover Publications.

Ericksen, S.C. (1941). The effect of distributed and massed practice on the nature of errors made by normal and operated rats. *Journal of Comparative Psychology*, *31*, 57–68.

Fu, Q., Bin, G., Dienes, Z., Fu, X., & Gao, X. (2013). Learning without consciously knowing: Evidence from event-related potentials in sequence learning. *Consciousness and Cognition*, *22*, 22–34.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*, 40.

Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, *29*, 831–848.

Glass, A. L., Brill, G., & Ingate, M. (2008). Combined online and in-class pretesting improves exam performance in general psychology. *Educational Psychology*, *28*, 483–504.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*, 212–242.

Hildreth, K. J., & Rovee-Collier, R. (1999). Decreases in the response latency to priming over the first year of life. *Developmental Psychobiology*, *35*, 276–290.

Izquierdo, I., Medina, J. H., Vianna, M. R. M., Izquierdo, L. A., & Barro, D. M. (1999). Separate mechanisms for short- and long-term memory. *Behavioural Brain Research*, *103*, 1–11.

Jin, I., Kandel, E.R., & Hawkins, R.D. (2011). Whereas short-term facilitation is presynaptic, intermediate-term facilitation involves both presynaptic and postsynaptic protein kinases and protein synthesis. *Learning & Memory*, *18*, 96–102.

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. III (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399–414.

McDaniel, M. A., Anderson, J. L., Derbish, M. H, & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494–513.

Naqib, F., Farah, C.A., Pack, C.C., & Sossin, W.S. (2011). The rates of protein synthesis and degradation account for the differential response of neurons to spaced and massed training protocols. *PLoS Computational Biology*, *7*, e1002324.

Packard M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neuropsychology of Learning and Memory*, *65*, 65–72.

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.

Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395.

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

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

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209–1224.

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, *63*, 87–92.
Rovee-Collier, C. K. (1993). The capacity for long-term memory in infancy. *Current Directions in Psychological Science*, *2*, 130–135.
Rovee-Collier, C., Evancio, S., & Earley, L. A. (1995). The time window hypothesis: Spacing effects . *Infant Behavior and Development*, *18*, 169–178.
Sadeh, T., Shohamy, D. Rubi Levy, D., Reggev, N., & Maril, A. (2011). Cooperation between the hippocampus and the striatum during episodic encoding. *Journal of Cognitive Neuroscience*, *23*, 1597–1608.
Seifert, L.S., & Baker, M.K. (2002). Art and Alzheimer-type dementia: A longitudinal study. *Clinical Gerontologist*, *26*, 3–15.
Seifert, L.S., & Baker, M.K. (2009). A report about current research: Maintenance of cognitive skills in persons with Alzheimer-type dementia. In L.S. Seifert (Ed.), *Roses grow in a butterfly garden* (pp. 199–238). Cuyahoga Falls, OH: Clove Press.
Solity, J. E. (2000). The early reading research: Applying psychology to classroom practice. *Educational and Child Psychology*, *17*, 46–65.
Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.
Squire, L. R. (1986). Mechanisms of memory. *Science*, *232*, 1612–1619.
Sumowski, J. F., Wood, H. G., Chiaravalloti, N., Wylie, G. R., Lengenfelder, J., & Deluca, J. (2010). Retrieval practice: A simple strategy for improving memory after traumatic brain injury. *Journal of the International Neuropsychological Society*, *16*, 1147–1150. doi: 10.1017/S1355617710001128
Sutton, M. A., Ide, J., Masters, S.E., & Carew, T.J. (2002). Interaction between amount and pattern of training in the induction of intermediate- and long-term memory for sensitization in aplysia. *Learning & Memory*, *9*, 29–40.
Underwood, B.J. (1961). Ten years of massed practice on distributed practice. *Psychological Review*, *68*, 229–247.
Viskontas, I. V., Carr, V. A, Engel, S. A., & Knowlton, B. J. (2009). The neural correlates of recollection: Hippocampal activation declines as episodic memory fades. *Hippocampus*, *19*, 265–272.
Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*, 464–476.
Zangwill, O. L. (1939). Some relations between reproducing and recognizing prose material. *British Journal of Psychology*, *29*, 370–382.

Original manuscript received December 1, 2012
Final version accepted April 15, 2013

Appendix

E. Which result indicates that a 4-year-old child may have a specific language disorder?
Poor performance at repeating letters
Poor performance at repeating words
Poor performance at repeating nonwords
All of the above
None of the above

R. One measure that predicts subsequent vocabulary growth for 4-year-olds is:
word recognition
word repetition
question comprehension
nonword repetition
word recall

E. Which of the following movie titles would be most memorable?
The Flea on the Dog
The Insect on the Animal
The Idea on the Subject
The Iff on the Ooff
124C41

R. Which college team name is most memorable?
Hokie
Hoya
Crimson
Chanticleer
Tiger