The Testing Effect and the Retention Interval

Questions and Answers

Thomas C. Toppino and Michael S. Cohen

Villanova University, PA

Abstract. People learn from tests. Providing tests often enhances retention more than additional study opportunities, but is this *testing effect* mediated by processes related to retrieval that are fundamentally different from study processes? Some previous studies have reported that testing enhances retention relative to additional studying, but only after a relatively long retention interval. To the extent that this interaction with retention interval dissociates the effects of studying and testing, it may provide crucial evidence for different underlying processes. However, these findings can be questioned because of methodological differences between the study and the test conditions. In two experiments, we eliminated or minimized the confounds that rendered the previous findings equivocal and still obtained the critical interaction. Our results strengthen the evidence for the involvement of different processes underlying the effects of studying and testing, and support the hypothesis that the testing effect is grounded in retrieval-related processes.

Keywords: memory, testing effect, cued recall, retention, recall, testing

Introduction

The *testing effect* occurs when adding tests to one's learning regimen benefits retention more than including a comparable amount of additional studying, particularly when the initial testing involves recall. This phenomenon has considerable potential for improving educational effectiveness (e.g., Dempster, 1992; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006a, 2006b), and understanding the underlying mechanisms should provide important insights into basic learning and memory processes (Dempster, 1996).

Because the to-be-remembered (TBR) items are reexperienced on a test, it is not surprising that testing enhances memory. Roediger and Karpicke (2006a), in a recent review, distinguished between mediated and direct effects of testing. In the former, testing facilitates performance indirectly by enhancing further study (e.g., processing of feedback and choice of study strategy). In the latter, taking a test per se facilitates performance (i.e., without additional study opportunities or feedback on the test). However, obtaining a direct effect of testing does not ensure that the underlying processes are fundamentally different from those involved in studying. Testing simply may be more effective at inducing essentially the same processes that are activated during study. Stronger evidence of qualitatively different underlying processes would be provided if a direct effect of testing could be reliably dissociated from the effect of studying.

Such a functional dissociation was reported by Wheeler, Ewers, and Buonanno (2003) who found that the type of learning experience (initial testing vs. additional studying of a list of nouns) interacted with retention interval in determining final free-recall performance. Studying yielded better retention than an initial free-recall test after a short retention interval, whereas initial testing was more beneficial after a long delay. Similar findings have been reported by others (e.g., Roediger & Karpicke, 2006b; Runquist, 1983). Wheeler et al. proposed that, while studying may strengthen a representation in memory, testing strengthens or enhances the retrieval process itself, reducing vulnerability to memory loss over time. Compatible views have been offered by others who have proposed that, in contrast to study trials, tests increase accessibility (Birnbaum & Eichner, 1971), provide retrieval practice (Runquist, 1983), or elaborate or increase the number of retrieval routes (Bjork, 1975). However, regardless of the hypothesized processes, the findings reported by Wheeler et al. are crucial. Unlike simple demonstrations of a testing benefit (i.e., a main effect), the interaction implies that testing and studying may affect retention in functionally different ways.

Although the existence of the testing effect and its potential for application seem unquestionable, the underlying mechanisms are less clear. Methodological concerns cloud the interpretation of many experiments in the testing effect literature, including the studies cited above. Many of the problems have been discussed by Carrier and Pashler (1992) and Kuo and Hirshman (1996).

One important issue raised by these authors is that differences in item presentation between initial tests and additional study opportunities often leave open the possibility that retention differences reflect mediated effects involving postretrieval study processes rather than the effect of retrieval processes per se. For example, item presentation during a test (especially free recall) is often self-paced, and this frequently is combined with output procedures that allow items to be viewed simultaneously. These practices provide opportunities for study and elaboration that are not as readily

DOI: 10.1027/1618-3169.56.4.252

available during study trials in which item presentation and exposure are more precisely controlled by the experimenter. In free recall, for example, self-paced output and the simultaneous availability of items during the test create an excellent opportunity to form inter-item associations that are especially important for performance on this kind of test.

Another critical issue concerns the comparability of the lists in the study and the test conditions of many experiments. Although participants are exposed to all TBR items on study trials, they typically recall only some of the items on test trials (Carrier & Pashler, 1992; Kuo & Hirshman, 1996). Thus, participants' exposure in the test condition is to an altered list that is often substantially shorter than the original list and contains only their best-learned items. This may redefine the list for participants in the testing condition, especially when multiple test trials are administered successively, as occurs frequently in the literature (e.g., Wheeler et al., 2003). Although participants in the study condition have an opportunity to study the very same well-learned items, they also are exposed to less well-learned items and whatever sources of interference the latter may introduce. The lack of list comparability in the study and the test conditions may affect final retention in at least two ways. In some cases, it may underestimate testing benefits because participants in the testing condition do not have an opportunity to study all the TBR items. Of greater concern, if a shorter list of well-learned items (testing condition) is retained better over time than a longer list that includes less well-learned items (study condition), the former may have an advantage at relatively long retention intervals. The combination of these effects might allow the critical interaction observed by Wheeler et al. (2003) and by others to be explained in terms of different study conditions rather than the beneficial effects of initial retrieval.

Thompson, Wenger, and Bartling (1978; Experiment 3) tried to maximize the initial recall by presenting and testing a series of five-item sublists. The final test on all words was given after a 10-min or 48-h retention interval. They found a small advantage of initial testing over additional studying after a long retention interval but not after a short one. However, Thompson et al. did not attain their goal of maximizing initial recall, and they exacerbated differences between the study and the test conditions by administering three repeated initial tests versus an equivalent number of additional study trials. Finally, their initial tests (written free recall) allowed simultaneous visual exposure to retrieved items during the test period, whereas items in the study condition received successive auditory presentation.

Kuo and Hirshman (1996) succeeded in minimizing procedural differences between study and test conditions, and virtually guaranteed successful initial retrieval. Each list item was presented and then was either restudied or tested after 3 s of distractor activity before the next item was presented. Following presentation of the last item, final free recall demonstrated superiority for tested items over restudied items. However, the implications of these findings are limited. Because initial recall was clearly mediated by immediate memory, one can question whether the results reflected the same processes as other findings in the literature that emphasized the effects of initial retrieval from long-term memory. Also, by using a single 5-min retention interval, Kuo and Hirshman could not assess how the effects of initial testing and restudying may have interacted with the length of the retention interval.

In a more recent study by Roediger and Karpicke (2006b, Experiment 1), participants studied a prose passage for 7 min and then either restudied it or engaged in written free recall during another 7-min interval. Following retention intervals of 5 min, 2 days, or a week, a final free-recall test was administered. The results indicated that restudying the passages produced superior final free recall after 5 min but that an initial free-recall test led to superior performance after the longer retention intervals. Although this study minimized many of the problems we described earlier, the results remain open to question because the free recall of prose passages may not provide adequate control over the way in which TBR information is processed. Roediger and Karpicke's participants were allowed a relatively long time period (7 min) during which they were free to restudy or recall in any manner they wished. There is no way to know exactly what strategies participants used during this interval, but it is at least possible that there were important differences between the restudy and the initial test conditions. For example, during the initial test, participants may have repeatedly reviewed what they had already recalled in an effort to spur further recall. Such repeated exposure to the idea units that were successfully recalled and that, by definition, were the most retrievable may have had different effects on subsequent retention than whatever strategies participants employed in the restudy condition.

It was, in part, the lack of sufficient experimental control in free-recall experiments that prompted Carrier and Pashler (1992) to suggest that cued recall of paired associates may be a more promising methodology for studying the testing effect. However, here too, results have been equivocal. Allen, Mahler, and Estes (1969) combined various numbers of study trials (cue and target presented together) with various numbers of test trials (only the cue presented). Test trials facilitated final cued recall after a 24-h retention interval, and error analyses suggested that studying leads to item storage whereas testing increases an item's retrievability, a conclusion similar to that proposed in the free-recall literature (e.g., Wheeler et al., 2003). Unfortunately, Carrier and Pashler argued that the results of Allen et al. could be questioned on grounds that the test trials were self-paced and fewer pairs were presented during testing trials than during study trials.

The interpretation of other paired-associate studies is also limited. Carrier and Pashler (1992) themselves obtained a testing effect, but the test trials were always followed by informative feedback, allowing the results to be interpreted as a mediated, rather than a direct, effect of testing. Runquist (1983) reported that testing reduces the rate of forgetting in paired associates, causing a greater testing advantage at longer retention intervals. However, his data are difficult to interpret. Initial testing was self-paced with all cues (and pairs, when retrieval was successful) being simultaneously present throughout the recall period, and results were reported primarily in terms of relative measures of retention. Still other studies (e.g., Carpenter & DeLosh, 2005; Cull, 2000) are limited by low levels of recall on the initial tests and/or by the use of only one short retention interval, precluding assessment of the interaction of learning condition (study vs. test) with retention interval that may constitute the strongest evidence that studying and testing effects are mediated by different processes.

The present research investigated the testing effect and especially its critical interaction with retention interval while eliminating or minimizing methodological concerns that have raised questions about previous experiments. In both the experiments, recall was compared for a study and a test condition after either a short (several minutes) or long (48 h) retention interval. In both experiments, we eliminated methodological problems previously associated with the use of free-recall and/or participant-paced initial tests by investigating the effects of learning condition and retention interval on vocabulary learning, using cued recall of paired associates and a single, experimenter-paced, initial cued-recall test. In Experiment 2, we also sought to minimize the difference between the lists to which the participants in the study and in the test conditions were exposed on their final learning trial (study trial or initial test) by providing them with an unusually high level of training prior to the final trial in order to ensure a high level of initial recall in the testing condition.

Experiment 1

The purpose of Experiment 1 was twofold. First, we wanted to assess the extent to which the methods introduced in this experiment could eliminate the critical interaction between learning condition and retention interval even though relatively low initial recall in the testing conditions would lead to functionally different lists being presented on the final acquisition trial for the study and the test conditions. Second, assuming that the interaction persisted, the results would provide a point of comparison for the results of Experiment 2. A concern about the high level of learning induced in the second experiment is that it would likely lead to very high final-recall performance after a retention interval of only several minutes, and this could complicate the interpretation of an interaction between learning condition and retention interval. Under such a circumstance, it would be useful to compare the interactions obtained in the two experiments.

Method

Participants and Design

Forty-eight introductory psychology students participated individually. Twelve were randomly assigned to each of the four groups generated by a 2 × 2 between-participants factorial design involving learning condition (study vs. test) and retention interval (2 min vs. 48 h).

Materials and Procedure

Lists contained 24 Swahili words paired with high-frequency English translations (Nelson & Dunlosky, 1994). Participants received four study trials in which pairs were presented successively in random order. Each pair was centered for 6 s on a computer screen with the Swahili cue above the English target. After four study trials, there was either an additional study trial or a test trial in which only cues appeared on the screen for 6 s each, and participants wrote their responses. After solving arithmetic problems for 2 min, half of the participants received the final cued-recall test, whereas the rest were dismissed. All participants were required to return in 48 h for a second session in order to guard against differential mortality rates in the two retention-interval conditions. Participants in the delayed-recall conditions received their final test in this second session. Final tests entailed a booklet with each page containing one Swahili cue and space for participants to write a response. Participants worked through the pages successively, at their own pace. The pages occurred in an independent random order for each participant.

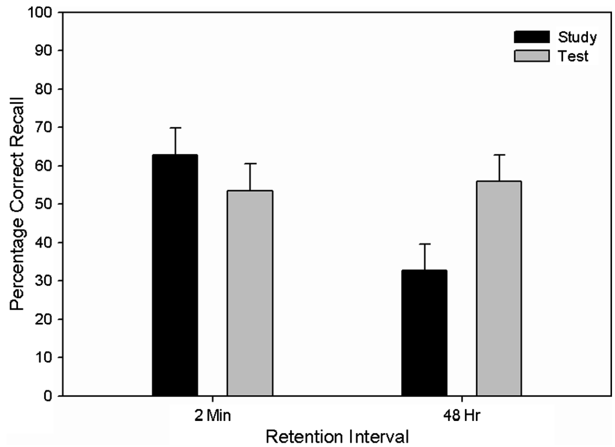

Figure 1. Mean percentage of correct cued recall on the final test as a function of learning condition (study vs. test) and retention interval (2 min vs. 48 h) in Experiment 1.

Results and Discussion

The data were the percentage of items correctly recalled on the final test by each participant. The analyses reported in this paper employed an alpha level of .05, unless otherwise specified.

Figure 1 shows the mean percentage of correct cued recall on the final test as a function of study condition (study vs. test) and retention interval. A 2 × 2 between-participant analysis of variance (ANOVA) revealed a marginal main effect of retention interval, $F(1, 44) = 3.93$, $MSE = 588.37$, $p < .06$, and a significant interaction, $F(1, 44) = 5.43$, $MSE = 588.37$. Simple effect analyses indicated that, although study superiority after 2 min was not reliable, $F(1, 44) < 1.00$, $MSE = 588.37$, test superiority after 48 h was, $F(1, 44) = 5.52$, $MSE = 588.37$. Thus, the pattern of results previously reported by others (e.g., Wheeler et al., 2003) was obtained when many of the methodological problems that have limited previous studies were avoided by using paired associates with a single experimenter-paced initial cued-recall test.

Experiment 2

Participants in Experiment 1, as in many previous studies, recalled only some (estimated to be a little more than half) of the TBR items on the initial test.[1] Thus, although the study and the test conditions involved learning the same nominal list, participants in the test condition may have studied a functionally shorter, better-learned list. Experiment 2 was a replication designed to increase list comparability by producing a high level of initial recall in the test condition. However, we elected to avoid training to a point that would yield 100% correct performance for fear that this would produce massive overlearning of many items that might obscure the very effect we were seeking. In our judgment, a small discrepancy in the number of items experienced on the final study and test trials was preferable.

Method

Participants and Design

The design was identical to that of Experiment 1. Sixty-eight introductory psychology students were assigned randomly to four conditions (17 per group). They participated individually or in groups of two or three. In addition, one participant failed to return for the second session and had to be replaced. A computer crashed on five occasions resulting in 12 participants being replaced (i.e., the participant whose computer crashed and the other participants in the room at the same time whose session was disrupted). Finally, there were four incidents of non-computer-related disruptions (e.g., insistent knocking at the door of the laboratory, incoming cell phone calls) resulting in five participants being replaced.

Materials and Procedure

The materials and procedures were the same as in Experiment 1, with the following exceptions. Participants sat before one of three workstations separated by dividers. Before the critical initial-test or additional-study trial, all participants received eight study trials on 20-pair lists. Pilot work indicated that this should yield at least 85% correct on the initial test trial. On all tests, participants had 6 s to type their response when each Swahili cue appeared on the monitor, but their responses did not show on the screen. The short retention interval was 5 min.

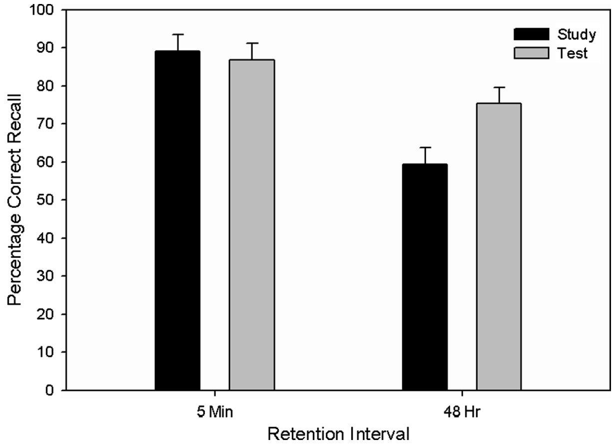

Figure 2. Mean percentage of correct cued recall on the final test as a function of learning condition (study vs. test) and retention interval (5 min vs. 48 h) in Experiment 2.

Results and Discussion

Final recall results are shown in Figure 2. A 2 × 2 ANOVA indicated a reliable effect of retention interval, $F(1, 64) = 22.41$, $MSE = 321.51$ and a significant interaction, $F(1, 64) = 4.40$, $MSE = 321.51$. Simple effect analyses indicated that, as in Experiment 1, additional studying produced a small, nonsignificant recall advantage after a short delay, $F(1, 64) < 1.00$, $MSE = 321.51$, whereas initial testing led to significantly better retention after a long interval, $F(1, 64) = 6.67$, $MSE = 321.51$. The high level of recall in Experiment 2 led us to compare the results of our two experiments in a 2 × 2 × 2 ANOVA with Experiment (1 vs. 2) as a factor. The critical learning condition by retention interval interaction was significant, $F(1, 108) = 10.58$, $MSE = 430.23$, but it did not vary across experiments despite marked differences in the level of recall in the two experiments, $F(1, 108) < 1.00$, $MSE = 430.23$, for the triple interaction. A similar outcome was obtained when the data from Experiment 2 were partitioned according to whether participants exhibited higher or lower levels of performance.

The present results replicated the pattern of findings obtained in Experiment 1 under conditions in which the level of initial recall was very high (84.4% and 85.9% for the short and long retention conditions, respectively). However, initial recall was not perfect. Could one, therefore, argue that the final-recall effects are attributable, not to the benefits of testing per se, but to differences in the number of items studied on the last study or test trial? We regard this possibility as highly implausible.

Participants in both the study and the test conditions studied all 20 list items for eight study trials before being exposed to a slightly different number on the final trial

[1] Due to researcher error, the initial recall data for Experiment 1 were unavailable for analysis. However, the level of initial recall performance should be similar to final recall after a short, 2-min retention interval.

(20 vs. 17 on average for the study and the test conditions, respectively). It seems highly unlikely that a very small difference in the number of studied items on one of nine trials would produce the obtained differences in retention. Something fundamentally different and dramatically more effective must have occurred on that last trial for it to overpower the extensive training that the study and the test conditions shared in common. We submit that the critical factor was that participants retrieved items in the testing condition.

General Discussion

Tests lead to better learning than additional study opportunities. This has been widely attributed to retrieval processes, with numerous theorists hypothesizing that retrieval practice during a test affects subsequent retention via mechanisms that differ from those operating in studying or encoding (e.g., Dempster, 1996). Crucial support for this theoretical conclusion lies in an interaction that dissociates studying and testing by showing that testing reduces the decline in performance as the retention interval increases. Thus, testing superiority is often observed only after a relatively long retention interval. However, as detailed in the Introduction, this critical interaction can be questioned on the basis of numerous methodological concerns.

The present experiments eliminated or minimized these concerns by investigating the retention of paired associates using a single, experimenter-paced initial cued-recall test and by varying the level of acquisition prior to the final learning trial involving an initial test or additional studying. In Experiment 1, a moderate level of acquisition led to modest initial recall, producing the kind of large discrepancy in exposure to list items on the final trial that has characterized previous studies. In Experiment 2, extensive acquisition training led to a high level of performance on the initial test, minimizing the difference between the study and the test conditions on the final trial. Yet, despite the marked differences in the level of performance, both experiments yielded the same crucial interaction: a nonsignificant advantage for the study condition after a retention interval of several minutes and a strong advantage for the test condition after 48 h. By confirming this crucial interaction while minimizing methodological concerns, these results provide strong evidence for the hypothesis that the effect of testing differs from that of studying because the former is mediated by processes that are unique to retrieval. Support for this position is magnified by the fact that previous supportive findings now appear unlikely to have resulted primarily from inadequate methodology.

Wheeler et al. (2003) suggested that, whereas studying may strengthen a representation in memory, testing may enhance the retrieval process itself. The benefits for retrieval have often been attributed to the strengthening of retrieval operations (e.g., Runquist, 1983) or the elaboration of retrieval routes (Bjork, 1975). An alternate, but not necessarily incompatible account of the testing effect, is provided by the concept of transfer-appropriate processing. In this view, memory is facilitated to the extent that the processes required for successful test performance match the encoding processes engaged during learning (Roediger & Karpicke, 2006a). Thus, when testing is part of one's learning experience, final test performance should be better than when it is not.

A recent study by Carpenter and DeLosh (2006) demonstrated that the benefit of initial testing on final test performance increases as initial retrieval is made more difficult: For example, by using increasingly impoverished cues. Furthermore, increasing the difficulty of initial retrieval benefited final test performance more than did matching the conditions of the initial and final tests. Carpenter and DeLosh concluded that elaborating and strengthening retrieval processes is a more important determiner of the testing effect than is transfer-appropriate processing. However, it should be noted that, although both hypotheses provide a possible explanation of why final test performance is better following an initial test than following an opportunity to restudy, neither necessarily explains why the difference between initial testing and restudying varies with the length of the retention interval. A transfer-appropriate processing account might suggest that the beneficial effect of matching test conditions with learning conditions is greater for longer retention intervals, while a retrieval strength account might suggest that retrieval processes that are strengthened and elaborated on an initial test are forgotten more slowly (Runquist, 1983). However, while plausible, both hypotheses largely redescribe the results and seem to require further development.

An interesting alternative by Bjork and Bjork (1992) was intended to explain both the testing effect and a number of other important memory phenomena. Bjork and Bjork distinguished an item's *retrieval strength* (current accessibility) from its *storage strength* (degree of learning). The probability of recall is attributed solely to retrieval strength. Storage strength does not affect performance directly but moderates changes in retrieval strength such that higher storage strength enhances increases in retrieval strength and slows the rate at which it is lost with the passage of time and other events.

Studying and retrieving information both are assumed to increase retrieval strength and storage strength, with the size of the increment in each being inversely related to the current level of retrieval strength. Thus, other things being equal, learning from successful retrievals is greater when retrieval is more difficult (i.e., when retrieval strength is lower), consistent with the findings of Carpenter and DeLosh (2006). Successful retrieval also is assumed to produce larger increments in retrieval strength and storage strength than restudying the same information. This assumption accounts for the superiority of initial testing over restudying, but the basis for the assumption, other than the finding it purports to explain, is not entirely clear. One possibility is that, to the extent that current retrieval strength depends on the current constellation of cues, the context of restudying information may maximize the current retrieval strength, reducing the size of the increment in retrieval strength in comparison with testing conditions. The interaction of type of learning experience (study vs. test) with retention interval is explained via two assumptions that were presented earlier. First, in comparison with restudying, successful retrieval not only produces larger increments in retrieval strength but also larger increments in storage strength. Second, the rate at which retrieval strength is lost over the retention interval is slower to the extent that storage strength is higher.

Although theory of Bjork and Bjork (1992) accounts for the basic phenomena associated with the testing effect, the theory needs further specification to clarify its relationship with other theories and to enable precise testable predictions. Nevertheless, the theory is attractive because it accounts for the testing effect in terms of assumptions that also provide a plausible account of other memory phenomena, including the recovery of older memories over time, the effect of overlearning on retention, and the spacing effect. With additional assumptions that were not germane in the present context, the theory also addresses retrieval-competition effects such as retrieval-induced forgetting. Thus, the theory places the testing effect in a conceptual context that suggests relationships with other important memory phenomena that might otherwise seem unrelated.

In closing, we note that recent testing-effect research has combined an interest in theoretical issues with a concern about educational applications. For example, Roediger and Karpicke (2006b) demonstrated that results similar to ours can be obtained with relatively complex textual material. Chan, McDermott, and Roediger (2006) showed that testing enhances retention for both tested and related, but untested, information. Our own research is consistent with this trend to the extent that we studied the testing effect in the context of vocabulary acquisition. However, the primary contribution of the present paper is that it helps to confirm that the empirical footing for both theoretical and practical interpretations of the testing effect is, in fact, as solid as the previous work has assumed it to be.

References

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning & Verbal Behavior, 8*, 463–470.

Birnbaum, I. M., & Eichner, J. T. (1971). Study versus test trials and long-term retention in free recall learning. *Journal of Verbal Learning & Verbal Behavior, 10*, 516–521.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 396–401). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus flucutation. In A. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619–636.

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

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

Chan, J. C. K., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553–571.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235.

Dempster, F. N. (1992). Using tests to promote learning: A neglected classroom resource. *Journal of Research and Development in Education, 25*, 213–217.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (pp. 317–344). San Diego, CA: Academic Press.

Kuo, T., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology, 109*, 451–464.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*, 325–335.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74*, 18–22.

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

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

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition, 11*, 641–650.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 210–221.

Wheeler, M. A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*, 571–580.

Received March 11, 2008
Revision received June 10, 2008
Accepted June 17, 2008

Thomas C. Toppino

Department of Psychology
Villanova University
800 Lancaster Avenue
Villanova
PA 19085
USA
Tel. +1 610 519 4722
E-mail thomas.toppino@villanova.edu

