BRIEF REPORT

Tests enhance retention and transfer of spatial learning

Shana K. Carpenter · Jonathan W. Kelly

Abstract Many studies have reported that tests are beneficial for learning (e.g., Roediger & Karpicke, 2006a). However, the majority of studies on the *testing effect* have been limited to a combination of relatively simple verbal tasks and final tests that assessed memory for the same material that had originally been tested. The present study explored whether testing is beneficial for complex spatial memory and whether these benefits hold for both retention and transfer. After encoding a three-dimensional layout of objects presented in a virtual environment, participants completed a judgment-of-relative-direction (JRD) task in which they imagined standing at one object, facing a second object, and pointed to a third object from the imagined perspective. Some participants completed this task by relying on memory for the previously encoded layout (i.e., the test conditions), whereas for others the location of the third object was identified ahead of time, so that retrieval was not required (i.e., the study condition). On a final test assessing their JRD performance, the participants who learned through test outperformed those who learned through study. This was true even when corrective feedback was not provided on the initial JRD task and when the final test assessed memory from vantage points that had never been practiced during the initial JRD.

Keywords Testing effect · Retrieval practice · Transfer · Navigation · Spatial learning

S. K. Carpenter (✉) · J. W. Kelly
Department of Psychology, Iowa State University,
W112 Lagomarcino Hall,
Ames, IA 50011, USA
e-mail: shacarp@iastate.edu

Retrieving information on a test benefits learning. Many studies have demonstrated this by presenting information to be learned (e.g., Swahili–English word pairs, such as *Farasi: Horse*) and then requiring participants to retrieve the information (e.g., Farasi: _____) or to spend an equivalent amount of time restudying it (e.g., Farasi: Horse). On a final test, information learned through retrieval is often remembered better than information learned through restudying (e.g., Roediger & Butler, 2011; see also Roediger & Karpicke, 2006a).

Commonly referred to as the *testing effect* or *retrieval practice*, this finding has been demonstrated across a wide range of verbal materials, including word lists (e.g., Carpenter & DeLosh, 2006), face–name pairs (e.g., Carpenter & DeLosh, 2005), paired associates (e.g., Carpenter, 2009, 2011), foreign language vocabulary (e.g., Carrier & Pashler, 1992; Finn & Roediger, 2011; Kang, 2010; Pyc & Rawson, 2010), general knowledge facts (e.g., Kornell, Hays, & Bjork, 2009; McDaniel & Fisher, 1991), and text passages (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler, 2010; Kang, McDermott, & Roediger, 2007; Roediger & Karpicke, 2006b). Demonstrations of the testing effect have been so numerous and reliable that it has been featured in a practice guide for educators as an instructional technique to improve student learning that is supported by strong evidence (e.g., Pashler et al., 2007).

Confidence in the educational benefits of testing is hindered by two major limitations, however. First, studies of the testing effect have been based almost exclusively on memory for verbal materials. Numerous demonstrations of the effect with word lists and paired associates have made no clear predictions about whether testing would enhance

more complex forms of learning that are not exclusively verbal, such as locations and distances within one's environment. Learning to navigate new environments is becoming increasingly important in military operations and transportation-based professions, however, so investigations of the potential benefits of testing for spatial learning would be of great practical value.

Only two known studies have explored the benefits of testing on learning two-dimensional map representations. Carpenter and Pashler (2007) presented maps that contained several features, such as roads, rivers, and trees. After having 20 s to encode the map, participants learned the features by having 100 additional seconds to view the map (i.e., the pure-study condition) or by spending 100 s viewing the same map with one feature randomly deleted at a time (i.e., the testing-with-feedback condition). In the latter condition, participants were informed that one of the features was missing and that they must try to remember the feature and its location. After mentally retrieving the feature, participants pressed a button that made the missing feature reappear so that they could score their accuracy. Thirty minutes later, when participants were asked to draw the maps, those maps that were learned through testing with feedback were drawn significantly more accurately than those that were learned through pure study.

More recently, Rohrer, Taylor, and Sholar (2010) had participants learn the locations of several cities on a map through testing with feedback versus restudying. In the testing-with-feedback condition, participants were given the name of one city at a time and were asked to place the city in its correct location on the unlabeled map. After each trial, participants saw the name of the city appear in its correct location. In the restudying condition, participants saw the name of each city appear in its correct location for the duration of the trial. On a final test in which participants were required to fill in the name of each city on an unlabeled map, the participants performed significantly better for maps that they had learned through testing with feedback rather than restudying.

These studies provide some encouragement that the benefits of testing may not be restricted to simple verbal stimuli. Additional data are needed, however, to explore whether testing is beneficial for more complex forms of spatial learning and to discover what may be driving these benefits. Both Carpenter and Pashler (2007) and Rohrer et al. (2010) provided feedback after the retrieval attempts, so the testing effects that they observed could have been due to retrieval itself, to better study allocation during feedback, or to some combination of these two factors.

In their review of the literature on testing effects, Roediger and Karpicke (2006a) distinguished between *direct* (i.e., retrieval-based) and *indirect* (i.e., feedback-based) benefits of testing, and this distinction has been important in guiding theoretical work on the testing effect in verbal-learning paradigms. For example, some hypotheses have sought to explain

how retrieval increases the effectiveness of encoding on a subsequent feedback trial (e.g., Izawa, 1992; Pyc & Rawson, 2010), whereas others have sought to identify aspects of the retrieval process itself that benefit learning even when feedback is not provided (e.g., Carpenter, 2009, 2011; Carpenter & DeLosh, 2006; Glover, 1989). Progress toward understanding the testing effect in any paradigm, therefore, would benefit by establishing whether the benefits of testing are direct or indirect. One of the objectives of the present study was to determine whether testing is beneficial for spatial learning, and furthermore, to establish whether this benefit should be characterized as direct or indirect.

The second major limitation of research on the testing effect is that in most studies a final test has been administered that measures memory for the same information that was retrieved on the intervening test. For example, after encoding a list of words, participants typically complete a final test that requires them to recall that same list of words. Although this has confirmed that testing is beneficial for direct retention of information, it does not inform us about whether testing enhances the *transfer* of learning.

In the map-learning study by Rohrer et al. (2010), significant testing effects occurred whether the final test required an activity similar to the intervening test (i.e., labeling the map by choosing city names from a list) or a different activity (i.e., recalling the city names before labeling the map, or naming a city that lies between two other cities). Carpenter and Pashler (2007) also observed significant testing effects on a final test that required participants to draw the maps, which was different from the computerized intervening test procedure that required retrieval of one feature at a time.

These studies indicated that testing is beneficial for learning maps, even when the final test measures memory for map content in a different way than had been tested previously. However, it has not yet been established whether testing benefits spatial representations that were *not* tested initially. For example, if participants learn the locations of points B and C by always using A as a vantage point, does this help them later when they must begin at C and find their way to B? Effective navigation is an important goal of spatial learning, and in many real-world situations individuals must reach a destination from a starting point that may differ from the one they have previously learned. Another objective of the present study, therefore, was to determine whether the act of testing spatial knowledge from one vantage point enhances later performance from a different vantage point.

In the present study, we explored the effects of testing versus restudying on retention and transfer of a complex three-dimensional spatial layout, and we sought to establish whether any benefits that occur from testing should be characterized as direct (i.e., resulting from retrieval) or indirect (i.e., resulting from feedback). Participants first encoded an array of familiar objects (e.g., a hat, car, and

plant) from a single view within an immersive virtual environment (see Fig. 1b). The array was then removed, and participants were asked to perform judgments of relative direction (JRD) in which they imagined standing at one object (e.g., the hat) while facing another object (e.g., the car) and indicated the direction of a third object (e.g., the plant).

Participants performed an initial JRD task in which they were required to complete several trials by adopting the same perspective as the view that they had originally encoded (i.e., the 0º perspective in Fig. 1a). Participants then performed a final test in which they were required to complete several JRD trials by adopting the same view as before (i.e., 0º) in addition to several new views that they had not previously experienced. For example, a trial from the 180º perspective required participants to imagine standing at the car facing the hat and to point to the ball (see, e.g., Fig. 1a). This design allowed us to measure direct retention of spatial knowledge (i.e., from the 0º perspective that was practiced), in addition to several measures of transfer based on perspectives that had never been practiced during learning.

During the initial JRD task, one group of participants was informed about the correct direction of the third object after they had tried to retrieve it (i.e., the test+feedback condition), and one group was not informed (i.e., the test-only condition). A third group of participants (i.e., study only) studied the same array and performed the same JRD task, except that this time the direction of the third object was always identified for them ahead of time, so that retrieval was not required. The final test phase was identical for all participants, and no feedback or visual indicators of the correct pointing direction were provided.

Method

Participants

A group of 64 undergraduates from Iowa State University participated in exchange for course credit. The data from 4 of these participants were removed due to average pointing errors that were worse than chance. The remaining 60 participants were randomly assigned to each of the three conditions, and participant gender was balanced across condition.

Stimuli

The virtual environment consisted of nine objects appearing on the ground of an infinitely large grassy plane (see Fig. 1b). These stimuli were viewed via a head-mounted display (HMD; nVisor SX111, NVIS, Reston, VA) on which binocular images of the virtual environment were presented at $1{,}280 \times 1{,}024$ pixel resolution within a 102º horizontal $\times$ 64º vertical field of view. The graphics presented via the HMD were updated at 60 Hz and reflected moment-to-moment changes in the participant's head position and orientation. Graphics were rendered using Vizard software (WorldViz, Santa Barbara, CA) running on a computer with Intel Core2 Quad processors and Nvidia GeForce GTX 285 graphics card.

Design and procedure

When participants donned the HMD, they were standing at the 0º view. The experimenter first named each object in a random order. Participants then studied the objects for 90 s, after which the objects disappeared. Following this encoding opportunity, participants removed the HMD and were led to another room to perform the JRD task.

To perform this task, participants were seated at a computer and given verbal instructions before beginning. Each JRD trial required participants to imagine standing at one object facing a second object, and to point toward a third object from the imagined perspective. The JRD task is illustrated in Fig. 2. The first object appeared in the center of a circle on the monitor, and the second object appeared at the top of the circle, thereby establishing the imagined perspective. The third object was listed at the bottom of the screen, away from the circle. The participants used a joystick to rotate a radial line emanating from the center of the circle until it pointed in the direction of the third object.

Fig. 1 Overhead (**a**) and perspective (**b**) views of the layout studied during the encoding phase
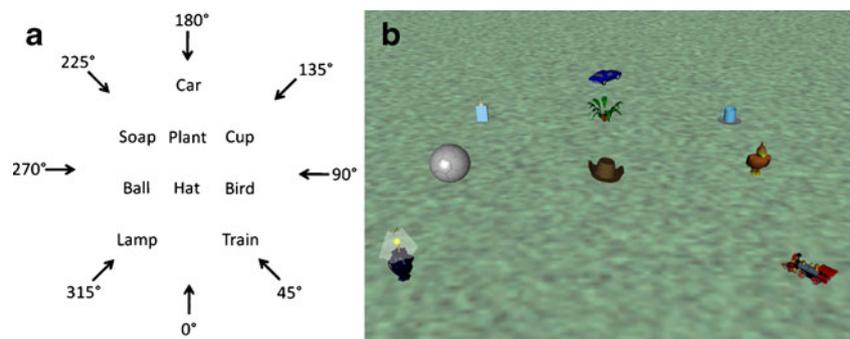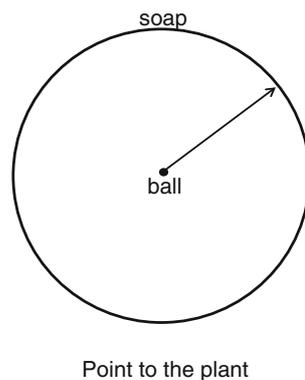
Fig. 2 Example of a judgment-of-relative-direction trial. In the example, the participant imagines standing at the ball while facing toward the soap, and must point to the plant from that perspective by manipulating the orientation of the radial line
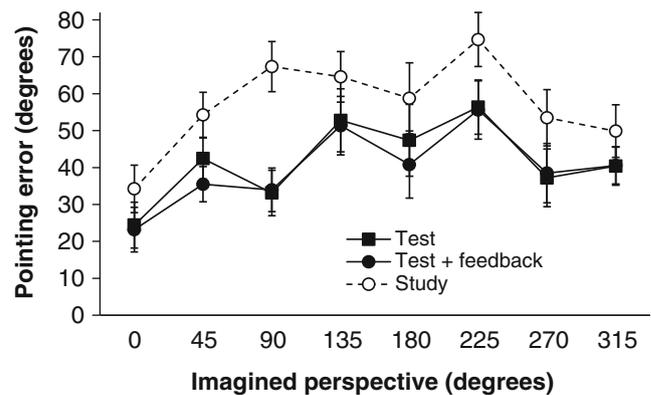
soap

ball

Point to the plant

Fig. 3 Absolute mean pointing errors as a function of imagined perspective and testing condition. Error bars represent ±1 standard error

A set of 6 unique JRDs were constructed from each of eight imagined perspectives, spaced every 45° from 0° to 315°, resulting in 48 total JRDs (see Fig. 1a). The initial JRD task comprised three repetitions of the 6 JRDs from the 0° perspective. Each participant experienced each of these 18 trials in a random order. For the final test, each of the 48 JRDs were presented once, in random order. The final-test trials assessing the 0° perspective were identical to those used during the initial JRD task.

During the initial JRD task, participants' experiences depended on the condition to which they were assigned. In the study condition, a marker on the circle indicated the correct pointing direction, and participants simply oriented the pointing line to that marker in order to advance the trial. In the test and the test+feedback conditions, participants pointed without such guidance. Responses in the test+feedback condition were followed by a marker indicating the actual direction of the object, which participants pointed to in order to advance the trial.

After completing the 18 trials of the initial JRD task, all participants completed a 10-min distractor task in which they answered a demographics questionnaire and spent the rest of the time trying to recall as many U.S. states as they could. Immediately afterward, all participants completed the same final test, consisting of the complete set of 48 JRD trials. On the JRD trials during the final test, no feedback or visual indicators of the correct pointing direction were provided.

Results

Figure 3 displays the mean absolute pointing errors during the final test across all imagined perspectives and conditions. A mixed-model analysis of variance (ANOVA) confirmed that the testing effect was significant, $F(2, 57) = 3.25$, $p = .046$, $\eta_p^2 = .10$. Pointing errors in the study condition ($M = 57.15°$, $SE = 5.25$) were significantly larger than those in the test condition ($M = 41.76°$, $SE = 5.25$), $F(1, 57) = 4.30$, $p = .043$, $\eta_p^2 = .07$, and in the test+feedback condition ($M = 39.92°$,

$SE = 5.25$), $F(1, 57) = 5.39$, $p = .024$, $\eta_p^2 = .09$, with no significant difference between the latter two conditions.

The main effect of imagined perspective was also significant, $F(7, 399) = 16.54$, $p < .001$, $\eta_p^2 = .23$. Performance was best when imagining the 0° perspective experienced during encoding ($M = 27.26°$, $SE = 3.46$) as compared to imagining all other perspectives ($M = 48.99°$, $SE = 3.16$), $F(1, 57) = 49.59$, $p < .001$, $\eta_p^2 = .47$. The interaction between perspective and condition was not significant.

The same analysis on final-test response times revealed a main effect of perspective, $F(7, 399) = 13.29$, $p < .001$, $\eta_p^2 = .19$. Participants responded faster when imagining the 0° perspective ($M = 13.4$ s, $SE = 0.56$) than when imagining all other perspectives ($M = 17.10$ s, $SE = 0.84$), $F(1, 57) = 41.65$, $p < .001$, $\eta_p^2 = .42$. Neither the main effect of condition nor the interaction between perspective and condition was significant.

Discussion

In the present study, we demonstrated significant benefits for spatial learning as a result of testing as compared to restudying. Consistent with the findings of a large number of studies on verbal learning (e.g., Roediger & Karpicke, 2006a) and two studies on map learning (Carpenter & Pashler, 2007; Rohrer et al., 2010), these results demonstrate that testing can benefit more complex forms of learning that are not easily characterized by verbal properties. The fact that the testing effect occurred even when feedback was not provided indicates that the retrieval process itself may help to strengthen spatial memory representations. Progress toward better understanding of the testing effect in spatial learning would therefore benefit by exploring aspects of the retrieval process that are likely to benefit learning.

One hypothesis that has been proposed to account for the direct benefits of retrieval is based on transfer-appropriate

processing (see, e.g., Morris, Bransford, & Franks, 1977), since an intervening test bears more resemblance to a final test than does an intervening restudy opportunity. The testing effect could therefore be due to the fact that the former is more likely than the latter to provide practice at the same type of activity that is required on the final test. According to this view, the testing effect should be stronger under conditions in which the intervening test and the final test are more similar, rather than different.

In the present study, we administered a final test over the same perspective that was practiced (i.e., 0º), as well as several new perspectives that were not practiced during learning (e.g., 90º and 180º). If the benefits of testing are strongest under conditions in which there is greater similarity between the intervening and final tests, the testing effect should be most pronounced on final test trials that require adopting the 0º perspective. Contrary to this prediction, we found that the benefits of testing applied across several different perspectives that had not been practiced during initial learning. Consistent with findings that spatial memories are orientation-dependent (e.g., Mou & McNamara, 2002; Shelton & McNamara, 1997, 2001), we found that participants were best at imagining the spatial layout from the encoding view, regardless of condition. At no time were the benefits of testing confined to a particular spatial orientation, however.

The benefits of testing on the transfer of learning are consistent with an increasing number of verbal-learning studies that have demonstrated significant testing effects under conditions in which the intervening and final tests were different (e.g., Butler, 2010; Carpenter, Pashler, & Vul, 2006; Chan, 2010; Chan, McDermott, & Roediger, 2006; McDaniel, Anderson, Derbish, & Morrisette, 2007). In a recent study, Kang, McDaniel, and Pashler (2011) demonstrated that testing can enhance transfer of mathematical function learning. Participants learned the relationship between x and y values either by seeing the two values side by side (i.e., pure study) or by attempting to estimate y given x before seeing the correct y value (i.e., test with feedback). On a final test, the participants demonstrated superior retention of previously learned y values if they had learned them through testing with feedback rather than through studying. Participants who learned the function through testing with feedback also performed better at estimating y values that were outside the range of those previously learned.

Other hypotheses that have sought to explain the direct benefits of retrieval have been based on changes in the organizational processing of the material that occur as a result of testing (e.g., Zaromb & Roediger, 2010). For example, in the semantic mediator hypothesis, Carpenter (2011) proposed that the act of retrieval activates semantic properties of a cue that could act as mediating information to facilitate later retrieval of the target. Although it is not clear

how the act of imagining a particular object in a spatial layout (e.g., a hat) would activate *semantic* attributes of that item (e.g., "head" and "hair"), retrieving the layout may activate other attributes—for instance, perceptual properties—that can mediate future retrieval. On a general level, therefore, the idea that retrieval improves the effectiveness of mediating information (see, e.g., Pyc & Rawson, 2010) is a viable hypothesis that could be applied toward understanding the testing effect in spatial learning.

The present design also helps rule out an artifactual explanation for the testing effect that is not usually addressed. Participants typically perform an activity during a test trial (e.g., typing an answer onto a computer screen or writing it down), whereas during a restudy trial they simply view the material again without performing this activity. This begs the question of whether testing effects could be due to the act of entering a response rather than to retrieval per se. In the present study, we controlled for this possibility by requiring participants in the study condition to perform the same activity as those in the test conditions—operating a controller device to indicate the direction of an object. Significant testing effects still emerged, indicating that these benefits are not likely to be due simply to the act of indicating a response.

In summary, the present study addresses the two greatest limitations in the vast research on testing effects. These data provide some encouraging news that tests can be effective for promoting both retention and transfer of complex spatial representations.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22,* 861–876. doi:10.1002/acp.1391

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1118–1133. doi:10.1037/a0019902

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569. doi:10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. doi:10.1037/a0024140

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.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14,* 474–478. doi:10.3758/BF03194092

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13,* 826–830.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642. doi:10.3758/BF03202713

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory, 18,* 49–57. doi:10.1080/09658210903405737

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571. doi:10.1037/0096-3445.135.4.553

Finn, B., & Roediger, H. L., III. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological Science, 22,* 781–786. doi:10.1177/0956797611407932

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

Izawa, C. (1992). Test trial contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy & S. M. Kosslyn (Eds.), *Essays in honor of William K. Estes: From learning theory to connectionist theory* (Vol. 1, pp. 1–33). Hillsdale, NJ: Erlbaum.

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition, 38,* 1009–1017. doi:10.3758/MC.38.8.1009

Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review, 18,* 998–1005. doi:10.3758/s13423-011-0113-x

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology, 19,* 528–558.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. doi:10.1037/a0015729

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.

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

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. doi:10.1016/S0022-5371(77)80016-9

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 162–170. doi:10.1037/0278-7393.28.1.162

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning* (NCER 2007-2004). Washington, DC: U. S. Department of Education, National Center for Education Research, Institute of Education Sciences. Retrieved from http://ncer.ed.gov

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335. doi:10.1126/science.1191465

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. doi:10.1016/j.tics.2010.09.003

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. doi:10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 233–239.

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review, 4,* 102–106.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43,* 274–310.

Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38,* 995–1008. doi:10.3758/MC.38.8.995