

Retrieval-Based Learning: A Perspective for Enhancing Meaningful Learning

Jeffrey D. Karpicke · Phillip J. Grimaldi

Published online: 4 August 2012
© Springer Science+Business Media, LLC 2012

Abstract Learning is often identified with the acquisition, encoding, or construction of new knowledge, while retrieval is often considered only a means of assessing knowledge, not a process that contributes to learning. Here, we make the case that retrieval is the key process for understanding and for promoting learning. We provide an overview of recent research showing that active retrieval enhances learning, and we highlight ways researchers have sought to extend research on active retrieval to meaningful learning—the learning of complex educational materials as assessed on measures of inference making and knowledge application. However, many students lack metacognitive awareness of the benefits of practicing active retrieval. We describe two approaches to addressing this problem: classroom quizzing and a computer-based learning program that guides students to practice retrieval. Retrieval processes must be considered in any analysis of learning, and incorporating retrieval into educational activities represents a powerful way to enhance learning.

Keywords Retrieval · Learning · Metacognition · Meaningful learning

What does it mean to say that a person has learned something? Some might say it means a learner has acquired new knowledge that now exists in his or her mind. Others might add that it means a learner has actively created a rich and elaborated knowledge structure. These ideas would tell only half the story, however, and it may not be the crucial half for understanding learning. Broadly defined, learning represents the ability to use past experiences in the service of the present. If a person has learned something, it means they are capable of using information available in a particular context, referred to as retrieval cues, to reconstruct knowledge in order to meet the demands of the present activity. Learning is therefore more than the encoding or construction of knowledge from experiences—it is the interaction between retrieval cues in the present and remnants of the past. Retrieval processes are involved in all situations in which knowledge is expressed, including situations

The writing of this paper was supported in part by grants from the National Science Foundation (DUE-0941170) and the Institute of Education Sciences in the US Department of Education (R305A110903). The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education. We thank Mindi Cogdill for assistance with manuscript preparation.

J. D. Karpicke (✉) · P. J. Grimaldi
Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette,
IN 47907-2081, USA
e-mail: karpicke@purdue.edu

where learners must produce the answer to a factual question, explain a concept, make an inference, apply knowledge to a new problem, and produce creative and innovative ideas. In all of those situations, learners draw upon the past in the service of the present; thus, all situations involve retrieval.

The perspective we refer to as *retrieval-based learning* is founded on two central ideas. The first idea is that retrieval is the key process for understanding learning and therefore must be considered in any analysis of learning. This is a fundamental idea in basic cognitive research on learning and memory, but the idea appears to be less influential in educational research. The second idea is that retrieval is not a neutral assessment of the contents of one's mind, but the process of retrieval itself contributes to learning. We describe several cases where retrieval influences learning both indirectly, by improving the quality of subsequent encoding, and directly, because the act of retrieval itself enhances learning. Importantly for educational applications, recent research has focused on using active retrieval to promote meaningful learning—learning of complex educational materials as assessed on measures of inference making and knowledge application. Despite the positive effects of active retrieval practice, several findings converge on the conclusion that many students lack metacognitive awareness of the benefits of active retrieval. In the final sections, we describe recent research on classroom quizzing and our recent efforts toward developing a computer-based system that guides students to practice active retrieval while they are learning.

Why Retrieval Is the Key to Understanding Learning

Many researchers working at the interface between educational and cognitive psychology—the present authors included—believe that such research should occur in “Pasteur’s quadrant,” according to the framework described by Stokes (1997; see Pintrich 2003). Stokes proposed that, rather than considering basic and applied research as a single continuum, the goals of research can be evaluated along two dimensions: the goal of deepening scientific understanding and the goal of promoting practical application. This can be represented as a 2×2 matrix in which research is relatively high or low in its quest for scientific understanding and relatively high or low in the quest for practical application. Stokes called the quadrant high in scientific understanding but low in practical application “Bohr’s quadrant” after Niels Bohr, the great physicist who contributed to fundamental understanding of atomic structure but who had little interest in practical application. Likewise, Stokes called the quadrant high in practical application but low in the goal of scientific understanding “Edison’s quadrant” after Thomas Edison, who was interested purely in practical application and not in scientific understanding. Pasteur’s quadrant is named for Louis Pasteur, who contributed both to basic understanding of microbiology and to solving practical problems of preventing disease. Research in Pasteur’s quadrant, therefore, would naturally integrate the quest for scientific understanding with the goal of practical application, because the goals are mutually informative. Focusing on applied problems can reveal new questions that deserve theoretical exploration, while theoretical ideas can be tools to aid in the development of new applications.

If we evaluate research at the intersection of educational and cognitive psychology according to the criterion of existing in Pasteur’s quadrant, we see two general challenges. One challenge is that research emerging from cognitive psychology, which may have implications for educational applications, is often not situated within educational contexts as firmly as it needs to be. In this article, we consider the case of active retrieval practice as an example of a finding from basic cognitive psychology with broader implications for

learning in education. We highlight ways researchers have attempted to integrate basic findings with educational materials and assessments by examining active retrieval in the context of meaningful learning.

The second challenge is that educational research should be tightly integrated with contemporary ideas about learning. Here we see another gap because, in many instances, educational research on learning has focused much more on encoding processes than on retrieval processes. Karpicke (2012) noted that the emphasis on encoding and the relative neglect of retrieval might be related to common metaphors people use to describe the mind and learning. Many metaphors used to describe the mind treat it as a place where knowledge is stored and treat knowledge as a collection of objects located in that storage space (Roediger 1980). Similarly, in education, a common metaphor invoked to describe learning is that of a physical building. Knowledge is *constructed* by learners who actively build knowledge *structures*; researchers seek to understand the *architecture* of the mind; and instructors aid students by providing *scaffolding* for learning (for discussions of this metaphor, see Palincsar 1998 and Stone 1998).

If the mind is viewed as a place where knowledge is encoded, constructed, and stored, then it is natural to focus on questions about the best ways to encode and construct knowledge in storage, and it is easy to see how processes involved in retrieving or reconstructing knowledge might not be viewed as central to understanding learning. Learning is often identified with the acquisition and encoding of new knowledge, and retrieval may be considered a separate matter involved in the measurement of the contents of mental storage. An assumption that stems from these views is that what a learner produces on an assessment reflects the contents of storage, or the knowledge constructed from previous encoding experiences, and failures to produce knowledge would be attributed to problems of encoding, storage, or construction processes. Neisser (1967) referred to this general view as the “reappearance hypothesis” because knowledge and experiences are encoded and assumed to reappear in the mind when they are needed. Likewise, Tulving (1974) referred to these as “trace-dependent theories” of learning because performance was thought to reflect the status of memory traces created during encoding.

However, for decades, researchers in cognitive psychology have made the argument on both logical and empirical grounds that people do not store copies of past experiences and reproduce them verbatim at the time of retrieval. This point is underscored by the fact that people sometimes experience illusions and distortions when they reconstruct knowledge. A tape recorder, video camera, filing cabinet, or any other device for recording, storing, and reproducing the past verbatim would not make such errors, but the human mind, which actively attempts to reconstruct the past, occasionally makes such errors as a consequence of its reconstructive nature.

More importantly, the knowledge a person expresses can vary greatly depending on the retrieval cues available in a particular context. In other words, we can hold all aspects of an initial learning experience constant—thus holding constant the conditions in which learners encode and construct knowledge—and still see wide variations in what people reconstruct during retrieval, based only on differences in the retrieval environment. Cognitive psychology research on learning and memory is replete with examples of how performance depends on retrieval conditions (e.g., see Tulving, 1983, chapters 10 and 11). An example with meaningful materials comes from classic research by Anderson and Pichert (1978). They had students read a meaningful story about events occurring in a house and then recall details from the story from one of two perspectives, the perspective of a burglar or the perspective of a person buying a home. After recalling the text from one perspective, the students then recalled again, but this time they were told to shift perspectives and recall

details of the story based on the other perspective. The key finding was that students recalled new information the second time that they had not recalled the first time. The newly recalled information was relevant to the perspective they adopted during the second recall period.

The only aspects that changed in the Anderson and Pichert (1978) study were the retrieval conditions. What then would one say about what students had “learned” from the experience of reading the story? The inferences one might make about student learning depend entirely on the retrieval conditions. The encoding and construction of knowledge during the original experience, and thus the traces of the experience established in the students’ minds, had been held constant across conditions, yet the knowledge students expressed changed dramatically when the retrieval cues and context were changed. The key lesson is that we can never directly examine what students have encoded, constructed, and stored from a learning experience. There is no way to create a mental vacuum and examine constructed knowledge outside of specific retrieval conditions. We can only examine what students are able to retrieve and reconstruct in a particular retrieval context with particular retrieval cues (Roediger 2000; Roediger and Guynn 1996).

It is for these reasons that understanding retrieval is essential for understanding learning, which is the first tenet of a retrieval-based perspective on learning. The second tenet, which we outline in the next section and elaborate in the remainder of this article, is that retrieval is not just a neutral assessment of a learner’s knowledge, but the act of retrieval itself produces learning.

The Process of Retrieval Influences Learning

Retrieval processes can influence learning in many ways, and our discussion is guided by an important distinction between indirect and direct effects of retrieval on learning (Roediger and Karpicke 2006a). An *indirect effect* of retrieval refers to situations where retrieval enhances learning by virtue of some other mediating process. For example, if a person attempts to retrieve knowledge, the outcome of the retrieval attempt would give the learner feedback that he or she could use to allocate study time or alter encoding strategies (Pyc and Rawson 2010). Another example is that instructors might encourage students to engage in retrieval by asking them questions in class, perhaps by using student response systems. This instructional strategy might motivate students to study more effectively in preparation for class participation. In both cases, retrieval would enhance learning indirectly by improving the processing that occurs when students encode knowledge.

In addition to these indirect effects, retrieval also produces *direct effects* on learning, because engaging in the process of retrieval itself produces learning. Every time we retrieve knowledge, that knowledge is altered, and the ability to reconstruct that knowledge again in the future is enhanced. A retrieval-based perspective on learning differs from a typical conceptualization of learning, in which learning is identified with the encoding of new knowledge and experiences and retrieval serves only as a means for assessing learning. Based on those ideas, increasing the frequency of encoding or study events would be expected to enhance learning, while increasing the frequency of retrieval opportunities would not be thought to affect learning, if retrieval simply involves the measurement of a person’s knowledge.

An experiment by Karpicke and Roediger (2008) directly examined those ideas about studying and retrieving knowledge. The experiment involved two sessions. In an initial learning session, students learned vocabulary words across a series of alternating study and retrieval periods. During study periods, students viewed a vocabulary word and its

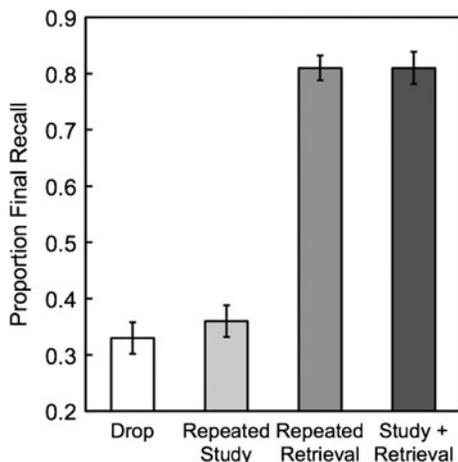
translation on a computer screen (*word* → *translation*), and during retrieval periods, they were given the vocabulary words as cues to recall the translations (*word* → ?). Students continued studying and recalling items until they had recalled each one in the learning phase. At that point, the vocabulary items were treated differently according to one of four different learning conditions. In a drop condition, once a word was correctly recalled, it was removed from further study and retrieval periods in the learning phase. In a repeated study condition, items were removed from repeated retrieval periods but were repeatedly studied. In a repeated retrieval condition, items were removed from repeated study periods but remained in retrieval periods. Finally, in a study-plus-retrieval condition, items were both repeatedly studied and retrieved. Thus, across the four conditions, students either repeatedly studied items or practiced repeatedly retrieving them during the learning phase.

The students returned to the laboratory 1 week after the learning phase and recalled the vocabulary words, and Fig. 1 shows final recall performance. If learning happens primarily when people study and encode experiences, we would expect to see large gains due to increasing the number of study opportunities. In contrast, repeatedly studying items produced no effect on retention. Likewise, if retrieval merely assessed the learning that occurred in prior study trials, then we would not expect to see much gained by increasing the number of repeated retrieval opportunities. Yet repeated retrieval produced large gains in long-term retention. Repeatedly retrieving words during initial learning, which amounted to only two or three extra retrievals in this experiment, produced about a 150 % improvement in long-term retention (see also Karpicke and Bauernschmidt 2011; Karpicke and Smith 2012).

Although the idea that retrieval produces learning may be counterintuitive to some readers, we imagine it may not seem surprising to all educational researchers. After all, educational psychologists have advocated generative learning activities for decades (e.g., Ausubel 1968; Wittrock 1974), and active retrieval practice may be essentially the same as other generative or active learning strategies already used in education. However, there is recent evidence from basic cognitive research that speaks directly to this issue and suggests that there is an important difference between simply generating knowledge and actively retrieving it.

A series of experiments by Karpicke and Zaromb (2010) directly compared the effects of generating words to the effects of actively retrieving words. Students viewed a list of target

Fig. 1 Proportion of vocabulary words recalled on a final recall test after a 1-week delay. Data adapted from Karpicke and Roediger (2008). Repeated retrieval during learning enhanced long-term retention, while repeated study of the vocabulary words produced no measurable effect

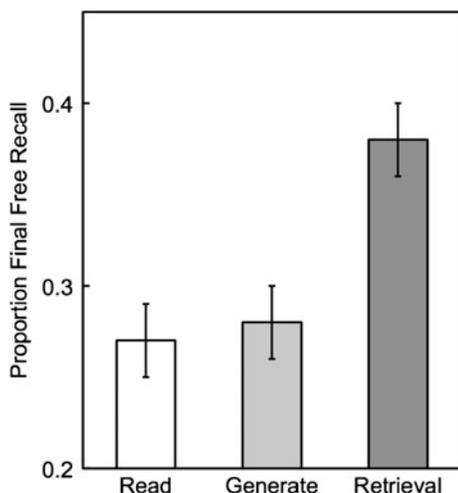


words (like *diet*) in an initial exposure phase. Then, in a second phase, students in a *read* condition studied the words paired with a cue word (e.g., *eat* □ *diet*), while students in two other conditions saw the cues paired with fragments of the target words (e.g., *eat* □ *di_*). In a *generate* condition, the students were told to generate the first word that came to mind that completed the fragment. In a *retrieval* condition, the students were told to use the fragment as a retrieval cue to recall a word they saw in the first part of the experiment. Therefore, the generate and retrieval conditions were exactly the same except for the instructions: Students in the retrieval condition were told to intentionally retrieve a word, while students in the generate condition were told to use a generation strategy. Importantly, performance in the initial generate or retrieval phase was identical in the two conditions (about 75 % correct).

The critical part of the experiment was a final free recall period a few minutes after the initial learning phase in which students were asked to recall all of the target items. Figure 2 shows the proportion of words recalled in the three conditions. The key result was that students in the retrieval condition recalled more items than did students in the generate condition. In fact, in this particular design, there was no advantage of generating over reading (i.e., no generation effect; see too Nairne *et al.* 1991; Slamecka and Katsaiti 1987). Therefore, this experiment represents a scenario where a generative learning activity produced no advantage over passive reading, but when the activity was converted into a retrieval-based learning activity, there was a significant effect. (Karpicke and Zaromb replicated this key result in three additional experiments.) This distinction between generating and deliberately retrieving knowledge has not yet been examined with more meaningful educational materials, to our knowledge, but the approach holds promise for enhancing the effectiveness of meaningful learning activities.

The previous studies show that active retrieval produces direct effects on learning but, as we have noted, retrieval can also produce important indirect effects on learning. In particular, one idea is that the act of attempting retrieval improves or “potentiates” encoding during a future study episode (see Izawa 1970; Karpicke 2009; Karpicke and Roediger 2007; Kornell *et al.* 2009). We recently carried out a series of experiments to examine the effects of attempting retrieval on subsequent encoding (Grimaldi and Karpicke 2012a). Students learned a list of word pairs (like *tide* □ *beach*), and, in what was called a “pretest” condition,

Fig. 2 Proportion of words recalled on a final recall test. Data adapted from Karpicke and Zaromb (2010). Intentionally retrieving words during the learning phase enhanced final recall, while generating words produced no advantage relative to reading words



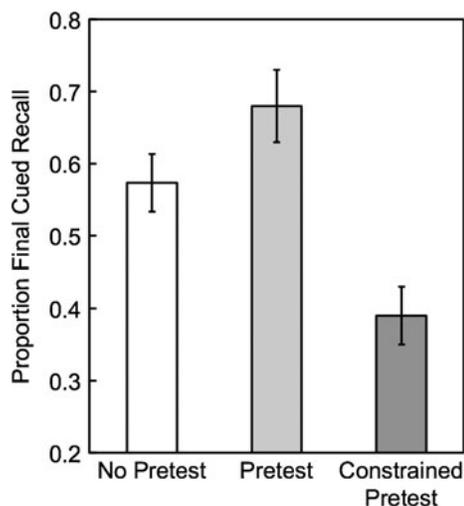
the students attempted to guess the target item prior to studying the pair (*tide*? prior to studying *tide*? *beach*). Because the target words were weakly associated with the retrieval cues, the students were unlikely to retrieve the target and typically retrieved a different word (for *tide*, the most frequent guess was *wave*). Performance in the pretest condition was compared to performance in two other conditions. In a no pretest condition, students studied the word pairs without attempting retrieval before each one. In a constrained pretest condition, students were given a fragment of a particular word to generate (e.g., *tide*? *wa* _ _), the target that students guessed most frequently).

The key measure was performance on an immediate final cued recall test of the target words, and Fig. 3 shows the proportion recalled on the criterial test. There are two key results. First, attempting retrieval enhanced performance (comparing the pretest to the no pretest condition). Second, the constrained pretest condition produced significantly worse performance than the no pretest control condition. This finding suggests that a broad search is more effective than a retrieval attempt that is constrained to a particular target. We argued that, during the process of a retrieval attempt, students establish a “search set” (Raaijmakers and Shiffrin 1981), and the act of specifying a search set during a retrieval attempt aids students when they subsequently encode new knowledge. Speaking more broadly, these results suggest that, attempting retrieval early in the process of learning, even before a person would be able to successfully recall desired knowledge, will help learners encode knowledge during study episodes (Karpicke 2009).

Practicing Retrieval Promotes Meaningful Learning

As we noted earlier, if cognitive psychologists wish to conduct educational research in Pasteur’s quadrant, a key challenge is to integrate cognitive principles into authentic educational activities. One might wish to dismiss the research described in the previous section because the studies used relatively simple materials and recall of word pairs as the measure of learning. Thus, it is clearly important to establish the effectiveness of retrieval-based learning activities with educational materials and assessments that reflect complex, meaningful learning.

Fig. 3 Proportion of target words recalled on a final cued recall test as a function of learning condition. Data adapted from Grimaldi and Karpicke (2012a). Attempting retrieval on a pretest enhanced learning relative to the no pretest control condition and constraining retrieval to a particular response on the pretest impaired learning



It is worth considering what is meant by the term “meaningful learning,” which is often defined in contrast to “rote learning” (Mayer 2008). Whereas rote learning is considered brittle and transient, meaningful learning is thought to be robust and enduring. Rote learning is thought to produce poorly organized knowledge, lacking coherence and integration, and is reflected in failures to make inferences and transfer knowledge to new problems. Meaningful learning, in contrast, is thought to produce organized, coherent, and integrated mental models that allow people to make inferences and apply knowledge.

However, as the preceding discussion of retrieval processes made clear, it is important to remember that, in all circumstances, people transfer past experiences to meet the demands of the present, and therefore, all circumstances involve reconstructing knowledge based on the cues available in a particular retrieval context. The situations thought to represent “rote” or “meaningful” learning may not reflect differences in what learners have encoded, stored, or constructed. Instead, the distinction between rote and meaningful learning depends upon the similarity of present retrieval scenarios to past learning experiences. Regardless of whether the goal of retrieval is to recall a fact, make an inference, or solve a new problem, the ability to accomplish the task depends on using retrieval cues to reconstruct knowledge.

Nevertheless, a wealth of recent research has extended active retrieval practice in two important ways. Active retrieval has been shown to enhance learning of meaningful educational materials, and it enhances learning on assessments designed to measure meaningful aspects of learning.

In one experiment using meaningful materials, Roediger and Karpicke (2006b) examined the effects of active retrieval practice on the learning of brief educational texts on science topics. Three groups of students read the texts under three different learning conditions. One group of students spent time repeatedly reading and studying the texts in four study periods (denoted SSSS). A second group read the texts in three study periods and then recalled the text in one retrieval period (denoted SSSR), in which the students wrote down as many of the ideas as they could recall from the text. A third group read the texts in one study period and then practiced recalling the text in three consecutive repeated retrieval periods (SRRR). Importantly, in this experiment, the students received no feedback after the retrieval periods; they purely practiced reconstructing their knowledge in the three consecutive retrieval periods.

At the end of this initial learning phase, the students were asked to predict how well they thought they would remember the material 1 week after the learning phase, a rating referred to as a *judgment of learning*. Then, 1 week later, the students returned to the laboratory and recalled the material again to see what they retained after the delay.

The right panel of Fig. 4 shows students’ judgments of learning and shows that their predictions were positively related to the number of times they repeatedly studied the material. In contrast, the left panel of Fig. 4 shows the proportion of ideas the students remembered 1 week after learning. Students’ actual long-term performance was positively related to the number of times they had practiced actively recalling during the learning phase. Even though all students spent the same amount of time learning the material, engaging in active retrieval practice produced substantially greater long-term retention than did repeated reading. It is worth mentioning that, although Roediger and Karpicke did not give students feedback after retrieval periods, other research has shown even larger effects of active retrieval practice when students briefly reread after attempting recall (see Karpicke and Roediger 2010). These data challenge the idea that retrieving and reconstructing knowledge is a “neutral” process. In fact, the data suggest that more learning occurred during repeated retrieval than during repeated encoding.

Several other studies have also shown that active retrieval enhances learning of educational materials (e.g., Agarwal *et al.* 2008; Butler and Roediger 2007; Kang *et al.* 2007;

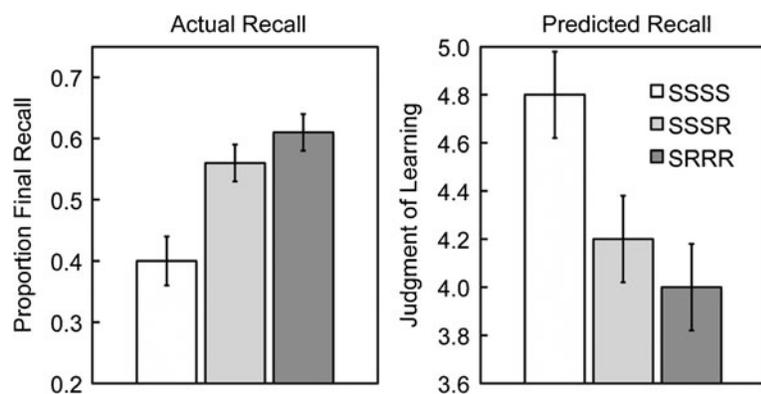


Fig. 4 Final recall (*left panel*) and judgments of learning (*right panel*) following repeated study or repeated retrieval practice of meaningful text materials. Data adapted from Experiment 2 of Roediger and Karpicke (2006b). The pattern of students' metacognitive judgments (predicted recall) was exactly the opposite of the pattern of students' actual long-term retention

McDaniel *et al.* 2009; among many others). Still, additional important questions remain if active retrieval is to be considered a viable approach to promoting meaningful learning. Specifically, it is essential to know whether active retrieval strategies are any more effective than other active learning strategies commonly used in educational contexts. It is also important to examine the effects of active retrieval practice on meaningful assessments of learning that measure the ability to make inferences, the ability to apply knowledge, and the coherence and integration of students' mental models.

Karpicke and Blunt (2011) took steps toward addressing these questions. They carried out two experiments designed to examine the effects of active retrieval practice on measures of meaningful learning and to compare the effects to those produced by another active strategy known as concept mapping (Novak and Gowin 1984; Novak 2005). Concept mapping is a popular activity that involves creating a diagram in which the concepts within some domain are represented as nodes and links connecting the nodes represent relations among the concepts. When students construct concept maps in the presence of the material they are learning, the activity bears the defining characteristics of an elaborative study method, because it requires students to enrich the material they are learning about by encoding meaningful relationships among concepts.

Karpicke and Blunt had students read educational texts and, in two key conditions, students either created a concept map of the concepts in the texts, or they practiced actively retrieving the ideas from the texts (using procedures similar to those in Karpicke and Roediger 2010, and Roediger and Karpicke 2006b). In two additional control conditions, students simply read the material either once or repeatedly. At the end of the initial learning phase, the students were asked to predict how much of the material they would remember in 1 week (as in Roediger and Karpicke 2006b). The students then took a final test 1 week later that involved two types of short-answer questions designed to assess meaningful, conceptual learning: *verbatim questions* that assessed conceptual knowledge directly included in the text and *inference questions* that required students to make connections across multiple concepts in the text.

The key results of the experiment were the proportions correct on a final short-answer test 1 week after the original learning phase, shown in Fig. 5a and b. The exact same pattern occurred for both question types. Elaborative studying with concept mapping produced only

a small gain in retention relative to reading once, whereas practicing retrieval produced the best long-term retention as assessed by both verbatim and inference questions. However, Fig. 5c shows the judgments of learning students made during the initial learning session, and the results show that students were not aware of the benefits of practicing active retrieval. In fact, students believed that repetitive reading would produce the best long-term learning and that elaborative concept mapping would be superior to practicing retrieval, even though retrieval practice produced the best long-term performance.

In a second experiment, Karpicke and Blunt again compared the effectiveness of creating concept maps or practicing retrieval during the learning phase. In this experiment, students created a concept map as the final assessment activity. Concept mapping has been used as an assessment of meaningful learning because it is thought to measure aspects of the coherence and integration of students' mental models of particular knowledge domains. Thus, concept mapping represents another tool to assess meaningful learning. As shown in the left panel of Fig. 6, active retrieval enhanced long-term learning even when the final assessment involved creating a concept map, an assessment that reflects the quality of students' conceptual understanding. The right panel of the figure shows students' judgments of learning and shows that students, once again, tended to believe that they had learned the material better after creating a concept map than after practicing retrieval. In sum, the two experiments by Karpicke and Blunt provide evidence that active retrieval practice was an effective strategy for learning meaningful materials, as assessed on measures of meaningful learning, and that a retrieval-based learning strategy was more effective than another active, elaborative study strategy.

The previous studies help establish that active retrieval is effective for learning meaningful educational materials as measured on meaningful assessments of learning. Earlier, we described experiments that have shown that attempting retrieval improves learning by potentiating subsequent encoding, because the retrieval attempt helps the learner establish a search set that facilitates the encoding of new knowledge (Grimaldi and Karpicke 2012a; Kornell *et al.* 2009). This finding with relatively simple laboratory materials has also been generalized to more complex, meaningful materials.

Richland, Kornell, and Kao (2009) recently examined the effects of attempting retrieval on subsequent learning of meaningful educational texts. Students in one condition answered pretest questions about concepts in the text prior to studying it, while students in a control condition only studied the text. Attempting retrieval by answering the pretest questions enhanced learning on final assessments both immediately after learning and after a week

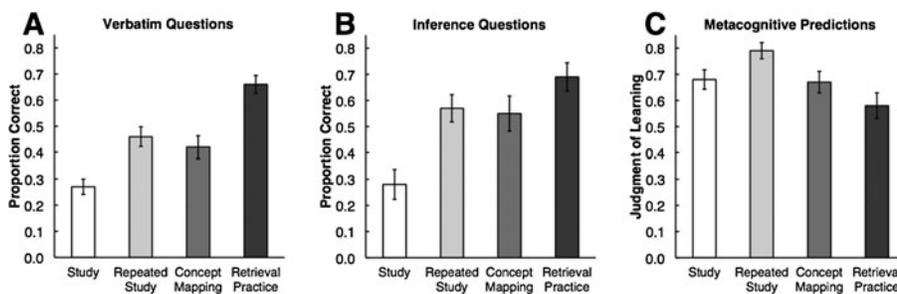


Fig. 5 Proportions correct on final short-answer verbatim questions (a) and inference questions (b) following a 1-week delay, and metacognitive judgments of learning during the initial learning phase (c). Data from Fig. 1 of Karpicke and Blunt (2011). Practicing retrieval enhanced long-term learning relative to elaborative studying with concept mapping, yet students were largely unaware of this benefit

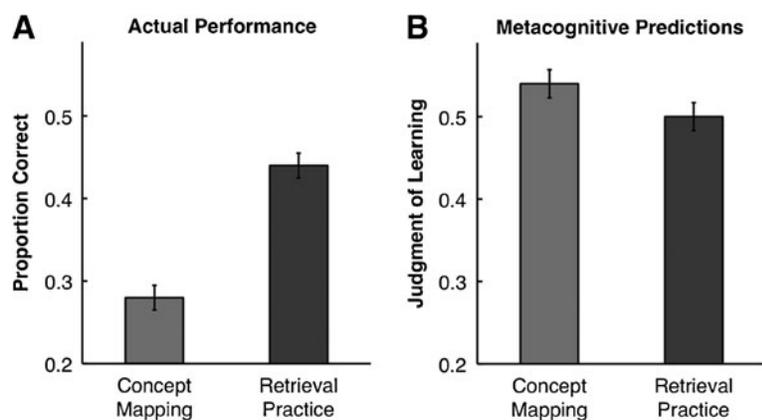


Fig. 6 Proportion correct on a final concept map test following a 1-week delay (a), and metacognitive judgments of learning during the initial learning phase (b). Data adapted from Fig. 2 of Karpicke and Blunt (2011). Practicing retrieval enhanced long-term learning relative to elaborative studying with concept mapping on a concept map test, yet students predicted that concept mapping would produce better learning

delay. Importantly, Richland *et al.* examined the idea that answering pretest questions simply signals readers to direct attention to particular portions of the texts. To test this hypothesis, Richland *et al.* gave students pretest questions and had them either attempt to retrieve an answer or simply study the questions without attempting retrieval. In both conditions, students were exposed to questions that would guide attention to particular portions of the texts, but the results showed that attempting retrieval of the pretest questions improved learning more than did merely reading the questions. These results support the idea that the process of attempting retrieval itself, rather than simply knowing particular relevant questions in advance, enhances subsequent encoding during reading. Thus, attempting retrieval, even before students are capable of producing correct answers, facilitates learning of complex, meaningful materials, probably because the retrieval attempt helps the learner establish a search set that facilitates subsequent encoding (Grimaldi and Karpicke 2012a).

Metacognition and Self-Regulated Retrieval Practice

If active retrieval is such a powerful tool for enhancing learning, it is worth asking whether students are aware of the benefits of retrieval practice and whether they use active retrieval practice when they regulate their own learning. The available evidence from laboratory studies and from surveys of students suggests that students are not generally aware of the beneficial effects of active retrieval for learning.

Several experiments have examined metacognitive awareness by asking students to make judgments of learning after they repeatedly read material, practiced retrieval, or engaged in other strategies, and many of these studies indicate that students are not aware of the positive effects of active retrieval. In the Karpicke and Roediger (2008) experiment described earlier, practicing repeated retrieval produced a large effect on long-term retention (see Fig. 1). When the students in the experiment were asked to predict how many items they would be able to recall in 1 week, they predicted they would recall about 50 % of the items, regardless of their learning condition. In the Roediger and Karpicke (2006b) experiment (Fig. 3),

students predicted they would recall more after repeatedly reading than they would after actively recalling material, even though the opposite was true. Karpicke and Blunt (2011) also found that students predicted they would remember more after repeatedly reading or after creating concept maps than they would after practicing retrieval (see Figs. 5 and 6).

Why do students fail to see the benefits of active retrieval relative to other encoding activities? Karpicke, Butler, and Roediger (2009) reasoned that students' judgments of learning are partially based on the fluency with which they process material (for review, see Koriat 2007). When students have material right in front of them, as they do when they repeatedly read, the material is immediately accessible and processing is fluent and easy. Thus, students' judgments of learning are high and generally overconfident relative to their actual level of performance on later assessments. In contrast, active retrieval changes the information students use to make their judgments of learning. Now, judgments are based on the ease (or difficulty) with which material can be brought to mind during retrieval.

Such metacognitive judgments have important consequences for the choices learners make when they regulate their own learning. Karpicke (2009) conducted a series of experiments examining students' self-regulated use of retrieval practice. Students studied and recalled vocabulary items across a series of alternating study and recall periods, in a procedure similar to that used by Karpicke and Roediger (2008). In one experiment, once the students had successfully recalled an item, the students first made a judgment of learning, in which they predicted the likelihood that they would recall the item in the future (on a 0 to 100 % scale). Then, the students made a self-regulated study choice, with three options: They could repeatedly recall the item, repeatedly study it, or remove it from practice. The students would take a final test on the items 1 week later, so they knew that they were studying for an upcoming criterial test.

The key finding was that once students could recall items during the learning phase, they tended to believe that they had learned those items well enough to recall them in the future, as reflected in relatively high judgments of learning (in the range of 70–80 %). This was especially true when items were recalled quickly and easily. Consequently, students typically chose to remove these items rather than practice recalling them (or restudy them). However, long-term recall of items that were not repeatedly practiced was about 35 %, whereas repeated retrieval produced levels of retention that were twice as good as that. Thus, when students retrieved items easily during learning, their metacognitive judgments of their own learning were high, and this led students to remove items from practice rather than practicing retrieval, even though active retrieval would enhance learning.

In another experiment, Karpicke (2009) examined when students would choose to begin practicing retrieval during the course of learning, an issue related to the potentiating effects of retrieval described earlier (Grimaldi and Karpicke 2012a; Kornell *et al.* 2009). Given that attempting retrieval potentiates subsequent encoding, it stands to reason that learners ought to begin attempting retrieval early during the course of learning, because such retrieval attempts will help learners establish search sets that will aid the encoding of new information. In one experiment, Karpicke (2009) had students make judgments of learning and strategy choices (to recall, restudy, or remove items) during study periods, rather than waiting until after the students had successfully recalled the items. Rather than attempting retrieval of all items early in learning, students tended to wait and attempt recall later in learning. As a consequence, the rate of learning was slower than it was when students attempted retrieval of all items throughout the course of learning, thus benefitting from the potentiating effects of retrieval on encoding.

These laboratory studies showing that students lack metacognitive awareness of the benefits of active retrieval also converge with surveys of students' learning strategies.

Karpicke *et al.* (2009) surveyed a large sample of undergraduate students by asking them to describe the strategies they used when they study and to rank-order their strategies. By far, the most frequently listed strategy was repeated reading of notes and textbooks: 84 % of students listed this as a strategy, and 55 % of them indicated that repeated reading was the #1 strategy they used to learn. In contrast, only 11 % of students indicated that they practiced actively retrieving or reconstructing knowledge while they studied, and just 2 % of students indicated that retrieval practice was their most frequent strategy. It is clear that practicing active reconstruction of knowledge is not something students typically do when they regulate their own learning.

Retrieval Practice in the Classroom

The research reviewed in the previous section suggests that many students do not recognize the benefits of practicing active retrieval and do not use it as a study strategy. The current challenges are to identify pedagogical strategies and develop new learning activities that involve retrieval practice. In the remaining sections of this article, we describe research on two attempts to meet these challenges: classroom quizzing methods and a computer-based guided retrieval practice program.

Several recent studies have shown that low- or no-stakes classroom quizzes, sometimes administered with clicker response systems, can serve as effective tools for promoting retrieval practice in the classroom (Mayer *et al.* 2009; McDaniel *et al.* 2011; Roediger *et al.* 2011; see too Campbell and Mayer 2009). McDaniel *et al.* (2011) conducted three experiments on the effects of quizzing with clicker response systems in middle school science classrooms. Students took multiple-choice quizzes covering material they were learning in different units in school (e.g., units on genetics, evolution, and anatomy). The quizzes occurred immediately prior to lectures on the topics, immediately after the lectures, and 1 day prior to the exams covering the units. Some of the concepts from within each unit were quizzed and some were not, and the effects of quizzing were assessed as performance on the unit exams (which occurred about 20 days after material was introduced in class) and on end-of-semester and end-of-year exams. McDaniel *et al.* observed positive effects of initial quizzing on all assessments, indicating that the benefits of engaging in retrieval in the classroom can be long-lasting. The results of these recent studies on classroom quizzing confirm that low-stakes quizzing is an effective way to get students to practice active retrieval in the classroom.

Some classroom studies have examined the effects of giving pretests or prequestions to students. The laboratory research reviewed earlier suggests that answering prequestions would benefit student learning but the benefits are likely to depend on the nature of the retrieval processes afforded by such pretests (Grimaldi and Karpicke 2012a). In one classroom experiment, McDaniel *et al.* (2011) manipulated whether students took multiple-choice quizzes prior to lectures, and those authors found no effects of taking pretests on learning. Other classroom studies have found benefits of pretesting. For example, Narloch, Garbin, and Turnage (2006) had college students in psychology courses answer fill-in-the-blank questions prior to lectures and found positive effects on students' classroom exam performance relative to a no-quiz control condition. Based on the theory that attempting retrieval on a pretest enhances learning when it helps students create a search set into which they can incorporate new knowledge (Grimaldi and Karpicke 2012a), it follows that different pretest formats might produce different effects on learning. Multiple-choice questions may not involve the same broad search processes as short-answer questions and thus

may not produce pretesting effects. It appears that most recent classroom experiments have examined multiple-choice quizzes, and to our knowledge, there has been no direct comparison of the effects of prequestion format on learning either in laboratory or classroom settings (for effects of the format of post-lecture quizzes, see Butler and Roediger 2007, and Kang *et al.* 2007). Further examination of pretesting effects is needed, and we predict that the effects will depend on the format of the prequestions and the nature of the retrieval processes afforded by the prequestions.

Finally, classroom quizzes do not need to involve clicker systems or take long amounts of time to benefit learning. Lyle and Crawford (2011) reported on a method they called “Practicing Unassisted Retrieval to Enhance Memory for Essential Material” (PUREMEM), which essentially involved giving students brief quizzes at the end of each lecture in a college statistics course. The nice features of Lyle and Crawford’s method were that the quizzes played a minor role in students’ grades and that the procedure emphasized to students the importance of quizzing to practice retrieval, rather than quizzing as a method of assessment. Not only did Lyle and Crawford find that quizzing enhanced student performance in class, but they also reported that students liked the frequent quizzing and viewed it as a valuable learning tool (see too Leeming 2002).

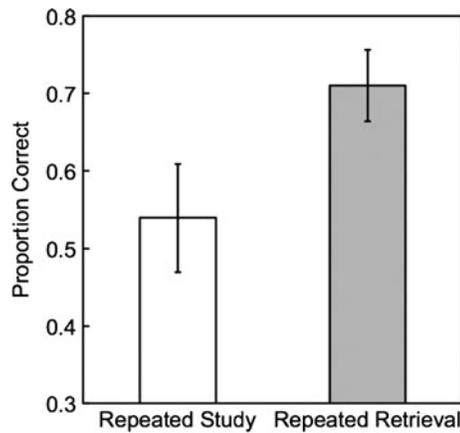
In concluding this section on classroom quizzing, we wish to emphasize that the fact that active retrieval promotes learning does not mean that instructors must include more tests and quizzes in the classroom. There are many ways instructors could incorporate active retrieval into classroom learning activities, and no single method has been identified as the best method. We suspect that group discussions, reciprocal teaching, and computer-based methods (explored in the next section) may be viable ways to encourage retrieval practice, but the role of retrieval processes in these activities has not yet been rigorously examined. We expect that future research will witness the development of novel and creative ways to integrate retrieval practice into educational activities.

Guided Retrieval Practice to Enhance Meaningful Learning

As noted earlier, many students do not use retrieval practice as a strategy when they study on their own (Karpicke *et al.* 2009), and when students regulate and control their study choices, they do not engage in retrieval as frequently as they should to produce the best learning (Karpicke 2009). To address these issues, we have recently been developing computer-based learning methods that involve *guided retrieval practice* (Grimaldi and Karpicke 2012b). During guided retrieval practice, students engage in repeated retrieval of complex educational materials, but study decisions are made by a computer instead of by the student. To accomplish this, we used natural language processing techniques to create a scoring algorithm capable of assessing students’ responses to short answer prompts. This algorithm, which we call *QuickScore*, scores students’ responses while they practice retrieval and determines when particular concepts would benefit from additional retrieval practice. Research on this approach is still in its early stages, so in this section, we describe a few preliminary studies using the QuickScore method to guide students to practice retrieval.

In one of our studies (Grimaldi and Karpicke 2012b), we used QuickScore to conceptually replicate Karpicke and Roediger’s (2008) retrieval practice procedure with more complex educational materials (human anatomy materials, specifically on the function, innervation, and location of particular muscles in the human body). Students studied and practiced retrieving these concepts across alternating study and retrieval periods. When QuickScore determined that a student had correctly recalled a concept, the concept was

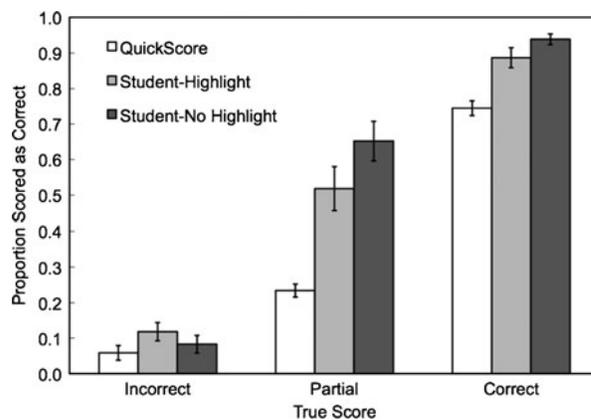
Fig. 7 Proportion of concepts correctly recalled after a 2-day delay. Data adapted from Grimaldi and Karpicke (2012b). Learning condition (repeated study vs. repeated retrieval) was manipulated using a natural language scoring algorithm, QuickScore. Repeated retrieval enhanced retention more than did repeated studying, as induced with the automated scoring algorithm



either presented in two repeated study periods (in a repeated study condition) or was recalled in two repeated retrieval periods (in the repeated retrieval condition). After a 2-day delay, students returned to the lab and attempted to retrieve the concepts again. Figure 7 shows the final results. Students in the repeated retrieval condition recalled significantly more concepts (about 20 % more) than did students who repeatedly studied the concepts the same number of times (Cohen’s $d=0.81$).

The results from the previous study show that guided retrieval practice with QuickScore is effective. However, does QuickScore actually score responses better than students do during self-regulated study? Moreover, can QuickScore be used to assist students in making better judgments of their own learning? In a follow up experiment, we directly compared QuickScore assessments to students’ own assessments of their performance during learning. We also examined the effects of using QuickScore to give students individualized feedback. Students practiced retrieval of anatomy concepts, using the procedure used in the previous experiment, except that immediately after each retrieval attempt the students compared their response to a model answer and scored it as incorrect, partially correct, or correct (see Dunlosky and Rawson 2012). After the students scored their own responses, the output from QuickScore was used to highlight key parts of the model answer that were missing from students’ responses. Performance in this highlight condition was compared to performance in a no-highlight control condition.

Fig. 8 Proportion of responses that were scored as correct by students in the experiment and by the QuickScore automated method as a function of the true score (scored by two independent raters). Data adapted from Grimaldi and Karpicke (2012b). Highlight indicates that missing key words were highlighted in the response correct response during the students’ self-scoring. Students were much more likely to give full credit to a partially correct response than was QuickScore



Students' self-score judgments and QuickScore's judgments were evaluated against the "true" score, as determined by two independent raters. The data in Fig. 8 show the proportion of responses counted as correct by the students and by QuickScore as a function of the true score (incorrect, partially correct, or completely correct). When a response was truly correct, students were more likely to rate it as correct than was the QuickScore program. Although students performed better than QuickScore in this regard, such scoring errors would likely benefit student learning, because during guided retrieval practice, QuickScore would simply keep these items in the practice list for longer amounts of time.

The crucial data in Fig. 8 are the responses that were only partially correct (according to the independent true score). If a partially correct response were scored as completely correct, the to-be-learned concept might be considered "learned" and might be removed from practice prematurely. This scenario would be detrimental to student learning. For partially correct responses, QuickScore was least likely to mistakenly score the response as completely correct (25 %). In contrast, students in the experiment gave themselves full credit for partially correct responses on 52 and 65 % of trials in the highlight and no-highlight conditions, respectively. Although highlighting portions of the correct answer had a positive effect by reducing the rate of erroneous self-scoring, the students' scores of partially correct responses were far less accurate than were the scores determined by QuickScore.

These preliminary studies suggest that guided retrieval practice is a promising approach to get students to engage in active retrieval. We hasten to add that the scoring algorithm we have developed in this guided retrieval approach is not completely new technology. Similar scoring algorithms and software already exist (e.g., C-rater; Leacock and Chodorow 2003), but these algorithms have been developed and used almost exclusively for assessment purposes, not as tools to promote learning by active retrieval practice. Our approach has been to bootstrap this automated scoring technology in a computer-based learning program that guides students to practice retrieval in the most effective ways.

Conclusion

Retrieval is more than merely an assessment of learning that occurred in a prior experience. The process of retrieval itself is central to creating learning, and it is always essential to consider the demands of a retrieval task in order to understand learning in any situation. The idea that active retrieval produces meaningful learning is broad and general, and there are likely many ways to integrate opportunities for active retrieval into particular learning activities. Thus, a retrieval-based perspective on learning recognizes that retrieval is a powerful tool for improving learning. The challenge for future research and development is to identify the best ways to leverage active retrieval to promote student learning.

References

- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*, 861–876.
- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecalled information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1–12.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart, and Winston.
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514–527.

- Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology*, *23*, 747–759.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271–280.
- Grimaldi, P. J., & Karpicke, J. D. (2012a). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*, 505–513.
- Grimaldi, P. J., & Karpicke, J. D. (2012b). *Guided retrieval of complex educational materials using computerized scoring*. Unpublished manuscript, Purdue University.
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired associate learning. *Journal of Experimental Psychology*, *83*, 340–344.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology*, *19*, 528–558.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*, 469–486.
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, *21*, 157–163.
- Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1250–1257.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772–775.
- Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.
- Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, *38*, 116–124.
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*, 17–29.
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*, 227–239.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*, 471–479.
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness* (pp. 289–325). New York: Cambridge University Press.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *35*, 989–998.
- Leacock, C., & Chodorow, M. (2003). C-rater: Automated scoring of short-answer questions. *Computers and the Humanities*, *37*, 389–405.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*, 210–212.
- Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*, 94–97.
- Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River: Pearson Merrill Prentice Hall.
- Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*, 51–57.
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, *20*, 516–522.
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399–414.
- Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *17*, 702–709.
- Narloch, R., Garbin, C. P., & Turnage, K. D. (2006). Benefits of prelecture quizzes. *Teaching of Psychology*, *33*, 109–112.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton.
- Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research In Science Education*, *35*, 23–40.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.

- Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh—A response to C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities". *Journal of Learning Disabilities, 31*, 370–373.
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667–686.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 333*, 335.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243–257.
- Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition, 8*, 231–246.
- Roediger, H. L. (2000). Why retrieval is the key process to understanding human memory. In E. Tulving (Ed.), *Memory, consciousness and the brain: The Tallinn conference* (pp. 52–75). Philadelphia: Psychology Press.
- Roediger, H. L., & Gynn, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 197–236). San Diego: Academic.
- Roediger, 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, H. L., & Karpicke, J. D. (2006b). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.
- Roediger, 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.
- Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26*, 589–607.
- Stokes, D. E. (1997). *Pasteur's Quadrant: Basic science and technological innovation*. Washington D.C.: Brookings Institution Press.
- Stone, C. A. (1998). The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities, 31*, 344–364.
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*, 74–82.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Witrock, M. C. (1974). Learning as a generative activity. *Educational Psychologist, 11*, 87–95.