2.27 Retrieval-Based Learning: A Decade of Progress

Jeffrey D Karpicke, Purdue University, West Lafayette, IN, United States

2.27.1 Introduction

Learning is often identified with the acquisition and encoding of new information. Reading a textbook, listening to a lecture, participating in a hands-on classroom activity, and studying a list of words in a laboratory experiment are all clear examples of learning events. Tests, on the other hand, are used to assess what was learned in a prior experience but are not typically viewed as learning events. The act of measuring knowledge—by recalling or recognizing items, by answering questions, or even by retrieving and applying knowledge to solve novel problems—is not thought to change knowledge, just as measuring one's height would not make one taller and measuring one's weight does not leave a person lighter.

Recent research in cognitive science has challenged the conventional view that tests only measure knowledge and are "neutral" for the process of learning. It may be obvious that tests can aid learning by providing feedback about what a person knows and does not know. But the more provocative result from recent research is that the act of taking a test—by itself and without any feedback or restudy—produces large effects on learning. This "testing effect" is driven by the retrieval processes that learners engage in when they take tests, and thus the key phenomenon is referred to as *retrieval-based learning*. Recent work on retrieval-based learning has led researchers and educators to rethink how learning happens and reappraise the role of testing in education.

Research on the role of retrieval in learning dates back a century (Abbott, 1909; Gates, 1917), and the topic has received occasional attention (for an historical review, see Roediger and Karpicke, 2006a). However, the past decade has witnessed a dramatic increase in research on retrieval practice. **Fig. 1** shows the number of papers between 1991 and 2015 retrieved from a search of Web of Science with the query "testing effect" or "retrieval practice" as a key term. The figure provides a clear picture of the explosion of research on retrieval-based learning over the past decade.

Several questions have guided this surge in recent research: What is the nature of retrieval-based learning? What are the mechanisms by which retrieval processes create learning? Do different initial retrieval practice conditions affect learning in different ways? Does retrieval practice enhance learning on a variety of outcome measures or only on specific types? Does retrieval enhance learning of all sorts of content, or do the effects differ for different materials? Do all people benefit from retrieval in the same way, or are there

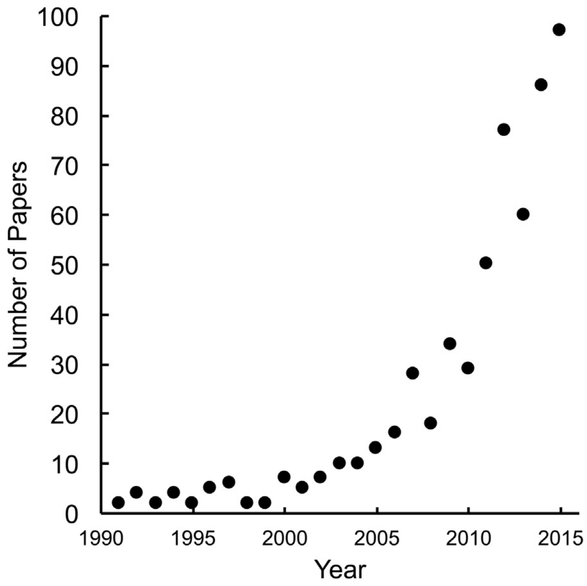

Figure 1 Number of papers on retrieval practice each year from 1991–2015, retrieved from a search of Web of Science.

individual characteristics that influence how much people learn when they practice retrieval? And do the benefits of retrieval practice generalize to educational settings and activities?

The present chapter contains several sections organized around these questions, focusing on new developments from the past decade of research on retrieval practice (roughly 2006–present). Following this introduction, the second section provides a primer on retrieval-based learning research, describing how researchers study retrieval practice effects and highlighting common methodological issues that arise in this research. The third section outlines four theoretical ideas that offer explanations of retrieval-based learning. The fourth section reviews a variety of factors that researchers have manipulated during initial retrieval practice activities, for the dual purposes of advancing theoretical understanding of retrieval-based learning and answering educationally relevant questions about retrieval practice.

Considerable recent attention has been paid to generalizing retrieval-based learning beyond laboratory settings and experiments with word-list materials. The fifth section reviews ways that researchers have assessed the effects of initial retrieval practice, especially with assessments that measure educationally meaningful learning outcomes. The sixth section reviews research that has examined how well retrieval-based learning generalizes across learner populations, to different types of materials, and to authentic educational contexts. The final section highlights a handful of open questions that await further research.

2.27.2 A Primer on Retrieval-Based Learning

The architecture of a retrieval practice experiment is fairly straightforward. In an initial learning phase, students study a set of materials. After studying, students in a retrieval practice condition complete one or more initial activities that require them to practice retrieving the materials. These activities are usually tests or quizzes, but learners can engage in a variety of activities that require them to retrieve knowledge but do not look like traditional "tests" (see Blunt and Karpicke, 2014). In a control condition, students do not practice retrieval. A variety of control conditions have been used in retrieval practice experiments, including having students complete no additional activity, spend extra time restudying the material, or complete other study activities (e.g., by engaging in an elaborative study activity). Several experiments have included more than one retrieval practice condition to compare the effects of different ways of practicing retrieval. Finally, students in all conditions take a final criterial assessment, which may occur anywhere from a few minutes after the initial activities (Rowland and DeLosh, 2015; Smith et al., 2013) to several months later (Carpenter et al., 2009; Larsen et al., 2013).

Although the general setup for a retrieval practice experiment is straightforward, a wide range of factors has been explored in retrieval-based learning research, including aspects of the initial retrieval practice conditions, the conditions of the final criterial assessment, characteristics of the learners, the nature of the materials, and the setting of the study (e.g., whether it occurred in a laboratory or a classroom). Before delving in to these factors, it is worth describing a handful of important methodological issues that investigators must bear in mind when they design and evaluate retrieval-based learning research, as detailed in the following sections.

2.27.2.1 Direct Versus Mediated Effects of Retrieval Practice

Retrieval can influence learning in a variety of ways. When students take a test (or test themselves), the outcome of the test provides information about what students know and do not know, and this information can guide future studying. Similarly, tests can

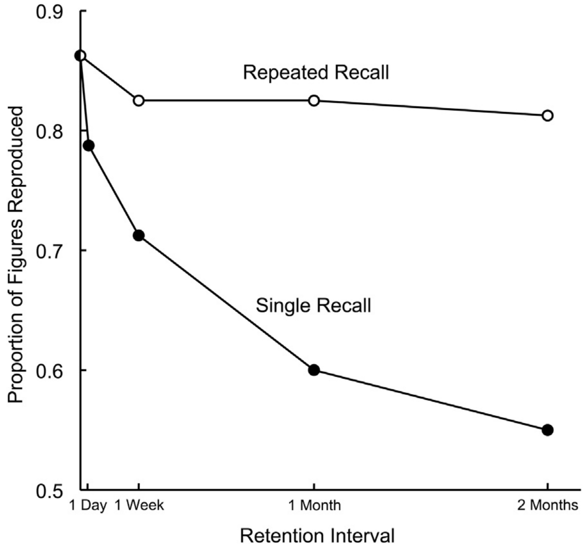

Figure 2 Proportion of line drawing figures correctly redrawn at varying retention intervals under repeated recall or single recall conditions. A typical forgetting trend is observed in the single recall condition, whereas little or no forgetting is seen in the repeated recall condition. Data are from Hanawalt, N.G., 1937. Memory trace for figures in recall and recognition. Arch. Psychol. 216, 89.

provide feedback to teachers that they can use to tailor their instruction, as in the practice of formative assessment (Black and Wiliam, 2009). Tests can also influence student motivation, because knowing about an upcoming test often leads students to increase their study efforts. Roediger and Karpicke (2006a) proposed that these are *mediated* effects of testing or retrieval on learning. In these scenarios, retrieval promotes learning by enhancing the processing that occurs during another activity (i.e., by making study activities more effective). Thus, the effect of retrieval on learning is mediated, for instance, by enhancements in subsequent restudy.

In addition to the mediated effects of retrieval on learning, retrieval also produces *direct* effects on learning. The direct effects of retrieval can be seen when students study a set of materials and then practice retrieval without restudying or receiving feedback after retrieval. Any gains in learning from practicing retrieval, without restudy or feedback, represent direct effects of retrieval processes on learning.

Although the present chapter is focused on recent research, an old experiment by Hanawalt (1937) provides an example of the direct benefits of repeated retrieval that is too remarkable to pass up. Hanawalt was interested in charting the curve of forgetting for people who recalled materials at only one point in time or repeatedly at multiple time intervals. He had subjects study a set of geometric line drawings and reproduce them at varying points in time: immediately after initial study, or 1 day, 1 week, 1 month, or 2 months later. Some groups of subjects recalled only once at one of the retention intervals. As shown in Fig. 2, when different subjects recalled the drawings once at different point in time, a typical pattern of forgetting occurred, with performance declining as the interval between the original study episode and the time of recall increased. A second group of subjects repeatedly recalled the drawings at each retention interval. Fig. 2 shows that in this repeated retrieval condition, there was little or no forgetting over time. Every time subjects practiced retrieving the line drawings, the act of retrieval enhanced the ability to retrieve the drawings again in the future.

A more recent example of the direct effects of retrieval practice comes from Roediger and Karpicke (2006b). One purpose of their experiments was to examine whether the benefits of retrieval practice would generalize to educationally relevant text materials. A second purpose was to compare retrieval practice conditions to repeated study conditions and ensure that total learning time was matched across conditions. In one experiment, students studied a brief educational text in one of three conditions. In one condition, the students repeatedly studied the text in a sequence of four study periods (labeled SSSS in Fig. 3). In a second condition, students studied the text in three study periods and recalled it once by writing down as much as they could remember (labeled SSST, where T indicates a free recall test). A third condition practiced repeated retrieval: students studied the text in one study period and recalled it three consecutive times (STTT). The important features of the experiment were that total time was matched across conditions and no feedback or restudy opportunities were provided following recall periods. Any effects would be due purely to differences in reading the texts versus freely recalling the texts.

The students then recalled the text again on final criterial test either 5 min or 1 week after the initial learning session, and the proportion of ideas recalled on the final test is shown in Fig. 3. When the final test occurred at the end of the experimental session, there was an advantage of having spent more time repeatedly studying the text. As discussed in more detail later, this advantage occurs because students reexperienced the entire set of materials in the repeated study condition, whereas they only experienced whatever they could recall in the repeated recall condition. The more important results occurred on the final test 1 week later: On this assessment of long-term retention, students who had repeatedly recalled the text three times remembered the most, more than students who had spent more time studying the materials in the other conditions. Repeatedly recalling the texts, without ever rereading, enhanced long-term learning.

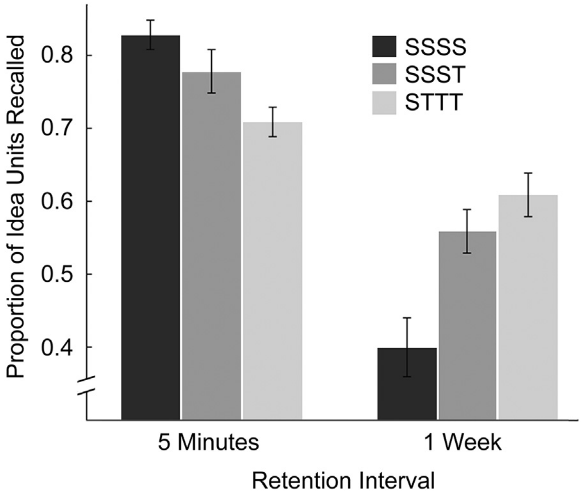

Figure 3 Proportion recalled on a final free recall test at either a 5 minute or 1 week delay by subjects who repeatedly studied (SSSS), studied and recalled once (SSST), or repeatedly recalled (STTT). Data are from Roediger, H.L., Karpicke, J.D., 2006b. Test-enhanced learning: taking memory tests improves long-term retention. Psychol. Sci. 17 (3), 249–255.

2.27.2.2 Balancing Retrieval Success and Retrieval Effort

Several of the methodological issues that crop up in retrieval practice research become apparent when considering the experiments just described. Hanawalt's (1937) results (Fig. 2) clearly show that repeated retrieval dramatically altered the shape of the forgetting curve. However, one could reasonably wonder whether the effects depicted in Fig. 2 are due to repeated retrieval, per se, or whether any reexposure to the line drawings would have produced similar results. That is, periodically restudying the line drawings may have produced effects similar to repeatedly retrieving them.

To alleviate that concern, many investigators have compared retrieval practice to conditions in which people restudy material, as Roediger and Karpicke (2006b) did. Several studies have shown that retrieval practice enhances retention even when compared to repeated studying, casting doubt on the idea that the results simply stem from reexposure to the materials. However, comparing repeated retrieval to repeated study introduces a new set of methodological concerns. Specifically, subjects in retrieval practice conditions typically do not successfully retrieve the entire set of to-be-learned material, while students in repeated study conditions reexperience all of the materials. There is almost always a difference in reexposure to the materials across conditions, a difference that depends on the level of *retrieval success* that students achieve in retrieval practice conditions.

Because of differences in reexposure to the materials across conditions, comparisons of retrieval practice and repeated study conditions are biased in favor of repeated studying. This makes it all the more impressive when retrieval practice outperforms repeated study, but it sometimes creates situations where no retrieval practice effect, or an advantage of repeated study, is observed. This pattern tends to occur when final tests are given shortly after the learning phase, leading some investigators to conclude that there are no benefits of retrieval practice at short delays and that the benefits only occur at long delays. However, this interpretation is incorrect, because the short-term advantages of repeated study are driven by differences in reexposure to the materials favoring restudy conditions. Indeed, several studies have observed retrieval practice effects at short delays (e.g., Karpicke and Zaromb, 2010; Rowland and DeLosh, 2015; Smith et al., 2013).

Ideally, one would want to maximize retrieval success to equate exposure across conditions. Some investigators have attempted to equate exposure by providing feedback on retrieval trials (e.g., see Carrier and Pashler, 1992). Although feedback does reexpose subjects to correct answers, and feedback is clearly beneficial when one's aim is to improve learning in educational settings, the provision of feedback introduces mediated effects and prohibits one from assessing the direct mnemonic effects of retrieval on learning. As an alternative, investigators might increase retrieval success by providing more initial retrieval support or by giving initial tests very shortly after material has been studied. However, conditions that increase initial retrieval success also make retrieval easier, and a good deal of evidence suggests that initial *retrieval effort* is an essential ingredient of retrieval-based learning. For example, providing more initial retrieval cues leads to smaller retrieval practice effects (Carpenter and DeLosh, 2006), and massed retrieval immediately after study produces very little learning relative to spaced retrieval (Karpicke and Roediger, 2007a). A later section on Manipulations of Initial Retrieval Practice Conditions reviews the evidence surrounding initial retrieval support. The conundrum for researchers, therefore, is how to create conditions that increase initial retrieval success without short-circuiting the benefits of initial retrieval effort.

Karpicke et al. (2014b) proposed two possible ways to bolster retrieval success without sacrificing retrieval effort. One approach is to design conditions that require retrieval effort (e.g., by ensuring that retrieval trials are spaced) and also afford a relatively high level of initial retrieval success. Based on Rowland's (2014) meta-analysis, retrieval practice effects become more robust as initial retrieval success increases, especially when initial retrieval is greater than 75%. Several studies have created conditions where initial retrieval is in this range and have shown reliable retrieval practice effects even when final tests occur in the short term (at the end of an experimental session; see Karpicke and Zaromb, 2010; Kornell et al., 2015; Rowland and DeLosh, 2015; Smith et al., 2013).

A second approach is to have subjects learn materials to the criterion of one correct recall of each item prior to manipulating repeated study or repeated retrieval of the items. Several articles have used this approach (Grimaldi and Karpicke, 2014; Karpicke, 2009; Karpicke and Roediger, 2007b, 2008; Pyc and Rawson, 2009). For example, Karpicke and Smith (2012) had subjects learn word pairs (e.g., Swahili–English vocabulary pairs) in a two-phase initial learning session. In the first phase, subjects studied and recalled the list across alternating study and recall blocks. Once each item was correctly recalled, it was removed from further blocks, and subjects continued this phase until they had recalled each item once (i.e., they learned the list to criterion). In the second phase, subjects either restudied the items or practiced retrieving them. Importantly, subjects successfully recalled approximately 85% of items during the repeated retrieval practice phase. Repeated retrieval enhanced subsequent retention of the items on a delayed test 1 week after the initial learning session. For present purposes, Karpicke and Smith's experiment illustrates how a learn-to-criterion procedure can help ensure high levels of initial retrieval success.

2.27.3 Theories of Retrieval-Based Learning

The fundamental question that theories of retrieval-based learning must address is why initial successful retrieval enhances the likelihood of subsequent retrieval, relative to control conditions in which learners do not practice retrieval. In Hanawalt's (1937) experiment, why did repeated retrieval essentially stop the curve of forgetting, allowing learners to maintain the same level of performance over long delays? Similarly, in Roediger and Karpicke's (2006b) experiment, why did recalling texts, without ever restudying or receiving performance feedback, produce better long-term retention relative to spending the same amount of time rereading? This section describes four theoretical ideas that have received the most attention during the past decade of research on retrieval practice.

2.27.3.1 Transfer-Appropriate Processing

Transfer-appropriate processing refers to the general idea that performance on a learning and memory task will be best when the processing that a student engages in during an initial learning activity matches or overlaps with the processing required during a later, final assessment (Kolers and Roediger, 1984; Morris et al., 1977). Because learners will be required to retrieve and use knowledge during a final criterial assessment, it makes sense that learners should practice retrieving and using knowledge during initial learning activities. Consider how people learn to play a musical instrument or to play a sport. If a person wants to learn how to play the violin, practicing by playing the instrument is essential; listening to someone else play the instrument, or reading about how to play it, is no substitute for practicing the skill itself. Likewise, learning to improve one's golf game, or to hit a baseball, or to perform well in perhaps any sport requires practicing skills, not merely watching others or reading about the sports.

As such, the idea of transfer-appropriate processing provides a useful heuristic for explaining why practicing retrieval during learning should bolster long-term retention. The importance of practice may seem obvious in other skill domains, but retrieval practice is not an obvious or widely used strategy in educational settings (see Karpicke et al., 2009). However, while transfer-appropriate processing may offer an intuitive explanation of the importance of retrieval for learning, it does not offer a mechanism that explains how or why engaging in retrieval improves learning. Moreover, a transfer-appropriate processing account leads to a prediction that has not been supported in retrieval practice research. A strict interpretation of transfer-appropriate processing is that performance will be best when the conditions and processing requirements on a final test exactly *match* the conditions and processing performed during initial learning. The greater the match between initial learning and final assessment conditions, the better the performance ought to be (cf. the transfer-appropriate processing account of implicit memory). However, several studies, reviewed later in this chapter, have shown that initial recall tests tend to promote greater gains in learning than initial recognition tests do, regardless of the format of the final assessment (recall or recognition). In sum, transfer-appropriate processing remains a useful heuristic for pointing up the importance of retrieval for learning, but other theories are required to explicate the mechanisms by which retrieval enhances learning.

2.27.3.2 Strength and Retrieval Effort

A second general idea about retrieval-based learning is the following: Practicing retrieval involves some effort on the part of the learner, and effortful retrieval of knowledge leaves that knowledge strengthened, increasing the likelihood that it can be accessed and used again in the future. Based on this general and appealing idea, retrieval practice represents a case of what Bjork and Bjork (2011) have referred to as creating "desirable difficulties" to enhance learning (see also Bjork, 1994, 1999). Bjork and Bjork have championed the idea that, rather than attempting to make activities especially easy for learners, activities that are difficult and require effort can be good for learning. Importantly, these conditions often lead to dissociations between initial learning and long-term performance (Soderstrom and Bjork, 2015). That is, certain conditions that make initial learning slower and more difficult may result in very good long-term retention and transfer; hence, those conditions constitute desirable difficulties. Conversely, conditions that lead to rapid initial learning may lead to very poor long-term retention and transfer. Unfortunately, such conditions would mislead students to think they are doing well, on the basis of rapid initial learning performance, even though their long-term performance may suffer.

Retrieval practice represents a desirable difficulty because it is more effortful to engage in retrieval than it is to use other strategies, like repetitive reading, yet retrieval practice confers great benefit in the long term. The crux of this view of retrieval practice is that the *effort* involved in retrieval is the key to learning (see Bjork, 1975). Specifically, effortful retrievals are assumed to strengthen retrieved knowledge, and stronger knowledge or memory traces are assumed to be more accessible in the future. The degree to which knowledge is strengthened is assumed to be proportional to the amount of effort involved in retrieval. More effortful retrieval conditions are assumed to produce greater gains relative to less-effortful retrieval conditions.

The notions outlined above were presented more formally by Bjork and Bjork (1992) in their new theory of disuse. That theory assumes that representations in memory have two strengths: a storage strength, which reflects the relatively permanent strength of a representation, and a retrieval strength, which represents its momentary accessibility at a particular time. Bjork and Bjork proposed that both storage and retrieval strength increase when an item is restudied, and both strengths increase to a greater degree when an item is retrieved. Importantly, Bjork and Bjork proposed that when an item with low retrieval strength is retrieved, it receives a greater increment in storage strength, relative to the increment in storage strength that occurs when an item with high retrieval strength is retrieved. Thus, effortful retrieval of items with low retrieval strength improves learning by producing large gains in storage strength.

An offshoot of these ideas about retrieval effort and strengthening deserves mention. Kornell et al. (2011) proposed a bifurcation account of retrieval practice specifically aimed at explaining situations in which retrieval practice effects depend on the retention interval (e.g., see **Fig. 3**; see also Halamish and Bjork, 2011). They proposed that initial recall tests produce a bifurcated distribution of item strengths, because some items are retrieved and consequently strengthened while others are not retrieved and are not strengthened. Similar to the ideas of Bjork and Bjork, the bifurcation account proposes that successfully retrieved items are strengthened to a greater degree than are items presented for restudy trials. The account assumes that the forgetting rate is the same for all items. Because all items in a restudy condition are strengthened a modest amount, while only the subset of recalled items are strengthened (but to a larger degree) in a retrieval practice condition, the study condition may outperform the retrieval practice condition on an immediate test. However, once forgetting has set in after a delay, the retrieval practice items, which accrued more strength initially, are more likely to be recalled than are restudied items.

There is good evidence that conditions that require effortful initial retrieval tend to confer greater benefits than do easier, less effortful retrieval conditions. However, the claim that effortful retrieval strengthens memory traces does not offer a mechanism to explain how or why effort would produce strengthening. The proposal that retrieved items become strengthened is essentially a redescription of the retrieval practice phenomenon instead of an explanation of it. Ultimately, the idea that effortful retrieval strengthens memory traces does not explain the deep structure of retrieval-based learning and, like transfer-appropriate processing, the account amounts to a heuristic description of how retrieval practice might work.

2.27.3.3 Elaborative Retrieval Account

The remaining two theories discussed in this section, the elaborative retrieval account and episodic context account, respectively, are relatively new theoretical ideas. Both accounts delineate possible mechanisms of retrieval-based learning, but the accounts differ in the nature of the underlying mechanisms they propose.

Carpenter (2009) proposed an elaborative retrieval account that has been expanded upon and developed in subsequent reports (Carpenter, 2011; Carpenter and Yeung, 2017; Kornell et al., 2015; Rawson et al., 2015a). The general idea is that semantic elaboration occurs during the process of retrieval and enhances subsequent recall. The concept of elaboration can mean a wide variety of things in memory research. In the elaborative retrieval account, elaboration is broadly conceptualized as the activation of cue-relevant information that becomes incorporated with successfully retrieved items (Carpenter and Yeung, 2017). More specifically, the idea is that when subjects are given a cue and asked to recall a target item, they generate several additional items that are semantically related to the cue. These items are incorporated with the target to form an elaborated memory trace that is memorable in the future.

A variety of evidence has been proposed to support the elaborative retrieval account, and this chapter reviews the evidence in detail. For the moment, one example from Carpenter (2009) will suffice. Carpenter had subjects study cue–target pairs where the target was either a strong semantic associate of the cue (e.g., *toast–bread*) or a weak semantic associate of the cue (e.g., *basket–bread*). Subjects then either restudied the pairs or took an initial cued recall test (e.g., they were given *toast–?* or *basket–?* respectively, as cues to recall the target *bread*). The assumptions were that when the cues were strong associates like *toast*, the target word would come to mind fairly readily during initial retrieval. On the contrary, for weakly associated cues like *basket*, subjects would need to generate several related words (like *eggs*, *fruit*, etc.) in an effort to recall the target. These additional verbal elaborations were assumed to be encoded with the target to form an elaborate representation that would be more robust and recallable in the future. Indeed, Carpenter found that initial recall with weak associates produced a larger retrieval practice effect than did initial recall with strong associates.

The elaborative retrieval account is appealing for a variety of reasons. It proposes that elaboration, which is thought to produce a variety of benefits for memory, also underlies the benefits of retrieval practice. The account also proposes a specific mechanism by which elaboration is thought to occur: During the process of retrieval, subjects activate several semantically related words that are then encoded along with the target to form a more elaborated and recallable representation.

However, the elaborative retrieval account has been placed under scrutiny in several recent articles (see Karpicke et al., 2014b; Lehman and Karpicke, 2016). The central idea of elaborative retrieval—that several cue-related items come to mind during retrieval—has been criticized on conceptual grounds. The idea seems incompatible with the operation of various global models of retrieval (e.g., Raaijmakers and Shiffrin, 1981). The idea also seems incompatible with retrieval-induced forgetting, which assumes that retrieval processes involve mechanisms that prohibit additional items from coming to mind, rather than facilitating the retrieval of several nontarget items. In addition, if many cue-related words did come to mind during retrieval, this ought to produce massive cue overload, making it more difficult to retrieve targets. Finally, attempts to directly induce the kind of elaboration proposed by the account have not produced results similar to retrieval practice effects (Lehman and Karpicke, 2016; Lehman et al., 2014).

Despite these potential shortcomings, the elaborative retrieval account enjoys popularity as an explanation of retrieval-based learning. The merits of the elaborative retrieval account are that it proposes a plausible mechanism for retrieval-based learning and it affords testable predictions (Lehman and Karpicke, 2016). The shortcomings of the account, however, leave open the possibility that other theories may be necessary to explain retrieval-based learning.

2.27.3.4 Episodic Context Account

Karpicke et al. (2014b) proposed an episodic context account of retrieval-based learning, which has also been developed and expanded upon in other articles (Lehman et al., 2014; Whiffen and Karpicke, 2017; see also Rowland, 2014; Rowland and DeLosh, 2014). The episodic context account explains retrieval practice effects on the basis of four central assumptions. First, people are assumed to encode information about items and the temporal/episodic context in which those items occurred (Howard and Kahana, 2002). Second, during retrieval, people attempt to reinstate the episodic context associated with an item as part of a memory search process (Lehman and Malmberg, 2013). The first two assumptions are grounded in models of memory and are not particularly controversial. Third, when an item is successfully retrieved, the context representation associated with that item is updated to include features of the original study context and features of the present test context. Finally, when people attempt to retrieve items again on a later test, the updated context representations aid in recovery of those items, and memory performance is improved.

An advantage of the context account is that attempts to provide a general account of retrieval-based learning based on principles from influential models of learning memory, namely the search of associative memory model (Raaijmakers and Shiffrin, 1981; see too Shiffrin and Steyvers, 1997) and temporal context model (Howard and Kahana, 2002), as well as ideas about how encoding variability occurs, to provide a general account of retrieval-based learning. A variety of evidence is consistent with the context theory. For instance, as discussed later, the context account offers explanations for why recall tests produce greater effects than recognition and why spaced retrieval is better than massed retrieval. Furthermore, manipulations of initial retrieval practice conditions have shown that requiring subjects to remember an initial episodic context enhances retrieval practice effects (e.g., Karpicke and Zaromb, 2010; Whiffen and Karpicke, 2017). Finally, several other studies have shown that initial retrieval practice enhances the ability to recollect episodic/contextual details on a final criterial test (e.g., Brewer et al., 2010; Chan and McDermott, 2007; Verkoeijen et al., 2011).

The elaborative retrieval and episodic context accounts can be seen as competing theoretical accounts that offer different views of the mechanisms that produce retrieval-based learning. However, as Carpenter and Yeung (2017) rightfully note, the accounts are not mutually exclusive. It may be the case that in certain circumstances or for certain types of materials, retrieval affords a great deal of semantic elaboration which in turn promotes retention, whereas in other circumstances retrieval conditions lead learners to think back to prior episodic contexts.

2.27.4 Manipulations of Initial Retrieval Practice Conditions

A considerable amount of recent research has manipulated aspects of initial retrieval practice activities for the dual purposes of exploring mechanisms of retrieval-based learning and identifying educationally relevant activities that enhance long-term learning. Initial retrieval practice conditions have been varied in several ways, but most research in this area coalesces on a common theme: Conditions that provide less retrieval support and require more effort from the learner tend to produce greater gains in learning, as long as learners can successfully retrieve material during the initial retrieval practice activities. Both elaborative retrieval and episodic context theories offer explanations for this general pattern of results. According to an elaborative retrieval account, conditions that provide less retrieval support and require more effort should afford more elaboration, and greater initial elaboration is assumed to promote long-term retention. According to the episodic context account, conditions that provide less retrieval support afford more context reinstatement, and greater degrees of context reinstatement and updating are assumed to enhance retention. This section reviews recent research on several ways that initial retrieval practice conditions have been manipulated, focusing on various ways researchers have manipulated the level of retrieval support during initial retrieval practice.

2.27.4.1 Retrieval Practice Compared to Restudy and Elaborative Study

Several studies have compared initial retrieval practice to conditions where learners restudy material, in an effort to equate the amount of time that people are exposed to materials across conditions. It is now well established that across a wide range of materials, initial retrieval activities, and final assessment conditions, retrieval practice enhances retention relative to repeated study. Rowland's (2014) meta-analysis captured this fact incisively. Across 159 studies, the overall effect size of retrieval practice relative to repeated study was $g = 0.50$, and 81% of comparisons favored retrieval practice over repeated study. It is beyond dispute that the benefits of retrieval practice cannot simply be attributed to reexposure to the material.

Recent work has compared retrieval practice to restudy conditions in which learners do not merely passively reread material but complete elaborative study activities that afford active engagement with the material. These studies have shown that retrieval practice tends to promote more learning than some elaborative study techniques do. In one study, Karpicke and Blunt (2011) had students read educational texts and either practice retrieving the material or create concept maps as an elaborative study task. Students in the retrieval practice conditions spent time writing as much as they could recall from the texts (following Roediger and Karpicke, 2006b). In concept map conditions, students drew node-and-link diagrams where nodes represented concepts in the text and links connecting the nodes represented relations among concepts (for an overview, see Karpicke, 2017). One week after the initial learning session, students took a final short-answer test that contained both verbatim questions focused on content stated directly in the material and inference questions that required learners to draw new connections among ideas. Retrieval practice consistently produced more learning than elaborative studying with concept mapping did (see also Lechuga et al., 2015).

Other studies have pitted retrieval practice against other forms of elaborative studying. In a series of experiments, Karpicke and Smith (2012) had students learn vocabulary words and showed advantages of repeated retrieval practice over elaborative imagery strategies, including the keyword mnemonic (Atkinson, 1975) and a mediator-generation strategy (Pyc and Rawson, 2010). Other studies have compared the effects of retrieval practice to the effects of elaborating on target words by generating several associated words that are related to the target word (Goossens et al., 2014b; Lehman and Karpicke, 2016; Lehman et al., 2014). These studies have also shown positive effects of retrieval practice relative to elaborative study control conditions. Collectively, research comparing retrieval practice to elaborative study provides strong evidence for the effectiveness of retrieval practice and poses a challenge to the idea that retrieval-based learning is driven by elaboration processes (Carpenter, 2009, 2011).

Another way to examine retrieval practice is to have students complete the same activity either while viewing the material, as an encoding task, or without viewing it, as a retrieval task. In educational settings, this distinction is often called open-book versus closed-book testing. Agarwal et al. (2008) directly compared open-book and closed-book quiz formats (see also Agarwal and Roediger, 2011). Students read educational texts and answered short-answer questions either with or without access to the texts. The open-book quiz conditions led to more forgetting over 1 week relative to the closed-book condition, in which learners attempted to retrieve answers before studying the answers. Thus, the same questioning or quizzing activity was made more effective when it required learners to engage in retrieval practice, rather than merely looking for answers in the text.

Finally, Blunt and Karpicke (2014) had students create concept maps either while studying texts, as an encoding activity, or without viewing the texts, as a retrieval practice activity. On a short-answer assessment 1 week after the learning session, students did better when they had learned by creating concept maps without viewing the texts, as a retrieval practice activity, than by creating maps while studying the texts. The concept mapping activity was made more effective when it required learners to engage in retrieval practice.

2.27.4.2 Comparisons of Recall, Recognition, and Initial Retrieval Cueing Conditions

Several experiments have varied initial retrieval cueing conditions as a means of manipulating the level of support during retrieval practice. These experiments often use lists of words or word pairs as materials to gain tight control over initial retrieval cueing conditions. The effects of initial free recall tests have been compared to the effects of initial item recognition tests, under the assumption that recall tests provide less support and require greater retrieval effort than recognition tests do. More recent work has introduced new ways to manipulate retrieval support by varying the nature of initial retrieval cues – for instance, by varying the number of letters in a word-fragment cue or varying the associative relatedness of cue–target word pairs.

Past experiments have compared the effects of initial recall to effects of initial item recognition tests on subsequent retention as assessed on final recall or recognition tests (e.g., Glover, 1989; Hogan and Kintsch, 1971). Somewhat surprisingly, direct comparison of initial recall and recognition testing has not received much attention in the past decade, but one exemplary recent study was reported by Carpenter and DeLosh (2006, Experiment 1). They had subjects study eight-item word lists and then either restudy the words or take an initial test, which was either a yes/no recognition test, a cued recall test where the first letter of each target was presented as a cue, or a free recall test. No feedback was given following the initial test. Subjects then took a final test, 5 min after the learning phase, in either recognition, cued recall, or free recall format. The results are shown in **Table 1**. Overall, as shown in the rightmost column of **Table 1**, initial free recall tended to produce the best performance on the final test, regardless of the final test format.

Carpenter and DeLosh's (2006) results support the idea that conditions with less initial retrieval support tend to promote better retention despite sizable differences in initial retrieval success across conditions, as shown in the left column of **Table 1**. For instance, whereas subjects correctly recognized 89% of the words, they freely recalled 69% of them, yet initial free recall consistently enhanced retention more than initial recognition did. Given such differences in initial retrieval success, researchers have sought ways to manipulate the level of initial retrieval support while minimizing any sacrifices to initial retrieval success. Carpenter and DeLosh (2006, Experiments 2 and 3) attempted to accomplish this by varying the number of letters of a target word that were presented during initial retrieval practice. Subjects studied five-letter words and then retrieved the words with four, three, two, or one letter of the word as cue. For instance, for the target word *cabin*, subjects received *c*_ _ _ _, *ca* _ _ _, *cab* _ _, or *cabi*_ as a retrieval cue, directly manipulating the amount of support provided during initial retrieval. Providing fewer letter cues during initial retrieval practice enhanced performance on a final free recall test.

Finley et al. (2011) varied the level of cue support across repeated retrieval trials of the same items. They had subjects study foreign language word pairs and restudy or practice retrieving under one of two conditions. In a diminishing cues condition, the

Table 1 Effects of initial retrieval practice under recognition, cued recall, or free recall conditions on final recognition, cued recall, or free recall performance

Initial test format		*Final test format*			
	Initial	*Recognition*	*Cued recall*	*Free recall*	*Overall*
Restudy	–	.56	.20	.28	.34
Recognition	.89	.53	.14	.20	.28
Cued recall	.72	.53	.16	.31	.33
Free recall	.69	.57	.29	.32	.39

Data from Carpenter, S.K., DeLosh, E.L., 2006. Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. Mem. Cogn. 34 (2), 268–276. http://dx.doi.org/10.3758/bf03193405.

number of letters of the target decreased across repeated retrieval trials (e.g., for the English–Inuit word pair *dust–apyuq*, subjects received the fragments *_p_uq*, *_p_u_*, and *_p_ _ _* across trials as retrieval cues). In an accumulating cues condition, the number of letters of the target increased across repeated trials. On a final cued recall test (e.g., *dust–?*), the diminishing cues condition consistently outperformed the study-only condition and the accumulating cue condition, which was no better than the study-only condition under conditions where subjects did not receive feedback.

Other work has varied the associative relatedness of cue–target pairs as a way to manipulate the amount of retrieval support provided by the task. Carpenter (2009) had people restudy or practice retrieval of word pairs that were strongly or weakly associated. The working assumption was that weak associates (e.g., *basket–bread*) offered less support and required more effort from the learner, relative to strong associates (e.g., *toast–bread*). On a final cued recall test, retrieval practice effects were larger with weak associates than they were with strong associates. Carpenter proposed that weak associates required subjects to engage in more elaborative retrieval relative to strong associates. The results can also be handled by the episodic context account, which proposes that weak associates require subjects to recollect prior occurrence information to a greater extent than do strong associates, which may come to mind more easily without recollecting an episode (Karpicke et al., 2014a).

In a similar vein, Carpenter and Yeung (2017) varied initial retrieval cueing conditions by using cue–target pairs that varied in "mediator strength," which refers to the degree to which a retrieval cue is associated with a nonpresented mediator word. For example, some word pairs had mediators that were strongly associated to the cues (e.g., for the pair *chalk–crayon*, the nonpresented mediator *board* is strongly related to the cue *chalk*). Other pairs had weakly associated mediators (e.g., for the pair *bomb–fire*, the mediator *explode* is weakly associated to the cue *bomb*). The assumption of the elaborative retrieval account is that production mediators is crucial for benefits of retrieval practice. Therefore, word pairs with strong mediators should produce larger retrieval practice effects relative to word pairs with weak mediators. Carpenter and Yeung reported that strong-mediator pairs were recalled better than low-mediator pairs, but this effect was not unique to retrieval practice – benefits of strong-mediator items were also seen in repeated study conditions. Although the data provide equivocal evidence for the elaborative retrieval account, the approach of examining mediator strength is an important and novel way to manipulate cueing conditions that deserves further exploration.

2.27.4.3 Retrieval Practice With Initial Short-Answer and Multiple-Choice Tests

A practical question with direct relevance to instruction is whether short-answer and multiple-choice question formats produce different effects on learning. From a purely practical standpoint, because multiple-choice questions are easier to score than short-answer questions are, it would be advantageous if multiple-choice questions enhanced learning to the same or to a greater extent than short-answer questions did. However, short-answer formats provide less retrieval support and presumably require more retrieval effort than multiple-choice formats do, and therefore, short-answer questions may produce more learning.

Research comparing the effectiveness of initial short-answer and multiple-choice questions has produced a decidedly murky picture. Some studies have shown advantages of initial multiple-choice questions, while others have shown advantages of initial short-answer questions, but feedback appears critical for observing benefits of initial short-answer formats. Still other studies have observed essentially no difference between short-answer and multiple-choice formats. Moreover, the original idea that short-answer questions require more effortful or recall-like processes than multiple-choice questions do has been challenged. Multiple-choice questions may require a great deal of retrieval effort or recollection, depending on how the multiple-choice alternatives are constructed (Little and Bjork, 2015; Little et al., 2012). Although research on short-answer and multiple-choice questioning dates back decades (e.g., Frase, 1968), the present chapter is focused on research from the past decade (for additional reviews, see McDermott et al., 2014; Smith and Karpicke, 2014).

Butler and Roediger (2007) had students watch videotaped lectures and take an initial short-answer or multiple-choice test. The initial short-answer and multiple-choice conditions used identical question stems; the only difference between conditions was whether learners produced answers to short-answer questions or selected an answer from several alternatives in multiple-choice conditions. On a final short-answer test 1 month later, performance was best when students had initially taken a short-answer test. Indeed, taking an initial multiple-choice test did not provide a benefit relative to a third condition in which students reread facts from the video. Interestingly, Butler and Roediger had students restudy the correct answers on half of the items on the initial tests, and they observed no effect of this feedback manipulation on the final test.

McDaniel et al. (2007a) compared the effects of short-answer and multiple-choice quizzes in an online college course. Students took Web-based quizzes, with feedback following the quizzes, or studied items in a restudy condition. The effects of quizzing or restudying were assessed on multiple-choice exams that occurred every 3 weeks and covered different units in the course. On the unit exams, both multiple-choice and short-answer formats produced positive effects, relative to rereading items, but initial short-answer quizzes produced larger effects than did initial multiple-choice quizzes.

Two critical factors in research on short-answer versus multiple-choice questioning are the level of initial retrieval success and the presence of feedback following initial questions. Initial short-answer and multiple-choice tests often produce different levels of initial retrieval success, just as initial recall and recognition tests do. For instance, students in Butler and Roediger's experiment correctly answered 68% of initial short-answer questions and 88% of initial multiple-choice questions. To combat differences in initial retrieval success, many experiments have included feedback following initial questions. This introduces mediated effects of retrieval practice (discussed earlier in the section Direct Versus Mediated Effects of Retrieval Practice), so it remains unclear whether the direct benefits of retrieval differ in short-answer or multiple-choice conditions. However, for practical purposes, this research helps address how much learning occurs when students take short-answer or multiple-choice quizzes that include feedback.

Kang et al. (2007) carried out two experiments in which they factorially crossed the format of an initial test (short-answer or multiple-choice) with the format of a final assessment (short-answer or multiple-choice). They had students read lengthy texts (2500 word articles), take an initial test, and complete a final test 3 days later. Students did much better on initial multiple-choice tests than they did on initial short-answer items (86% vs. 54% correct in their Experiment 1). It is therefore not surprising that students in the initial multiple-choice test did better on the final test than students who took an initial short-answer test did. This advantage occurred regardless of the format of the final test. Importantly, in a second experiment, Kang et al. gave students feedback following the tests by having them study the correct answers. In this experiment, students who took initial short-answer tests performed better than did students who took initial multiple-choice tests. Again, the advantage of initial short-answer tests occurred regardless of the final test format.

Little et al. (2012) carried out experiments similar to those of Kang et al. and obtained similar results. They found that when students took initial tests without feedback, taking an initial multiple-choice test produced better final test performance than taking an initial short-answer test did. However, when students received feedback, there was a slight advantage of initial short-answer testing over initial multiple-choice testing. Therefore, it may be true that answering initial short-answer questions affords greater retrieval effort and thereby produces greater gains in learning relative to answering initial multiple-choice questions. But because initial retrieval success is much greater on initial multiple-choice tests, the mnemonic benefits of short-answer tests are not observed unless students receive feedback.

However, more recent results have clouded this simple picture. Even when feedback is provided, some studies have not observed differences between short-answer and multiple-choice formats. McDaniel et al. (2012) carried out two experiments in an online course (similar to McDaniel et al., 2007b) in which students took short-answer or multiple-choice quizzes with feedback. On criterial unit exams, both quiz conditions outperformed a reread control condition, but there was no difference between short-answer and multiple-choice formats. McDaniel et al.'s study differs from other studies in several ways. Specifically, students could take quizzes up to four times, and the number of times students took quizzes was not controlled; students could take the quizzes at varying times, up to 1 h before the criterial exam; and the quizzes were open book, so students could use course materials while taking the quizzes (see Agarwal et al., 2008). Nevertheless, in this naturalistic setting there was no difference in the effectiveness of short-answer and multiple-choice quizzes.

McDermott et al. (2014) carried out a series of controlled experiments in a seventh grade science classroom. Students took either short-answer or multiple-choice quizzes in their classrooms throughout the semester. Short-answer responses were made on paper and multiple-choice responses were made using classroom clicker systems (see the later section on Educational Contexts), and students received feedback on all quiz questions. On unit exams and end-of-semester exams, both quiz conditions outperformed a no-test control condition, but there were either no differences between test formats or a slight advantage of multiple-choice over short-answer format.

Finally, Smith and Karpicke (2014) reported a series of laboratory experiments comparing initial short-answer and multiple-choice formats. They also examined a hybrid format, in which students first attempted to recall answers in short-answer format and then selected an answer from among alternatives in multiple-choice format. This hybrid format has been proposed as a possible way to combine the benefits of short-answer and multiple-choice formats (Park, 2005) but has not received much attention in experimental work. After completing the initial tests, students read correct answers as feedback. One week later, the students took final short-answer test that included both verbatim and inference questions (Karpicke and Blunt, 2011). All test formats produced robust effects relative to study-only control condition, but there were no discernable differences among the formats. Even conditions designed to bolster initial short-answer performance led to only modest enhancements in long-term retention, relative to multiple-choice testing. Across four experiments, Smith and Karpicke reported an overall effect size of $d = 0.07$ for the effect of short-answer versus multiple-choice questioning on final retention.

In sum, while there once seemed to be strong evidence favoring short-answer formats over multiple-choice formats, more recent data have led to a decidedly mixed set of evidence. There are often large differences in retrieval success in initial short-answer and multiple-choice conditions. Examination of direct effects of retrieval practice across different formats is clouded by differences in initial retrieval success. Including feedback ensures that students reexperience all correct answers but also introduces mediated effects. It has been assumed that short-answer questions would lead to more learning than multiple-choice questions because the short-answer format requires learners to engage in more retrieval effort. One reason that studies have not seen differences between test formats may be because multiple-choice questions do afford more retrieval effort (or elaborative retrieval, or context reinstatement) than originally assumed. Indeed, multiple-choice questions might be constructed to promote a greater effort, elaboration, or context reinstatement, as reviewed in the next section.

2.27.4.4 Positive and Negative Effects of Initial Multiple-Choice Questions

Research examining the effects of multiple-choice questions on learning has focused on two general topics. One topic is the potential negative effects multiple-choice tests, which may occur because learners are exposed to incorrect information (the lures on the test) that they might retain in the long term. A second topic is concerned with how multiple-choice questions might be constructed to afford greater levels of initial retrieval effort and thereby promote more learning (for review, see Marsh and Cantor, 2014).

Multiple-choice questions, by design, present learners with erroneous information: On each question, a correct response is embedded among multiple incorrect responses. An important question is what happens when learners select the wrong response on a test. In a series of experiments, Roediger and Marsh (2005) had students read texts and answer initial multiple-choice questions with varying numbers of alternatives (e.g., 2, 4, or 6 multiple-choice alternatives). Increasing the number of alternatives increased the likelihood of selecting an incorrect answer on the initial test. Importantly, on a subsequent cued recall test, students tended to produce the incorrect responses that they had selected on the initial multiple-choice test. The effects of the number of multiple-choice alternatives on subsequent retention depend on how successful students are on the initial tests (Butler et al., 2006). When the level of initial retrieval success is high, having more alternatives benefits subsequent retention. Presumably, increasing the number of multiple-choice alternatives makes initial retrieval practice more effortful, which benefits subsequent retention. However, when initial retrieval success is low, having more alternatives hurts learning. In this case, students are more likely to select incorrect responses on the initial test, and those incorrect responses are remembered on future tests.

Other work has explored implications of the negative effects of multiple-choice questioning. Marsh et al. (2009) had college students answer multiple-choice questions from SAT II tests on biology, chemistry, and history. Five minutes after completing the initial test, the students took a final short-answer test. There was a benefit of the initial multiple-choice condition relative to a no-test control condition (i.e., an overall effect of retrieval practice). However, taking the initial multiple-choice tests increased the likelihood that students would intrude incorrect items on the final short-answer test, relative to the no-test control condition. The implication is that the standardized tests taken by millions of students every year may create false knowledge, simply by exposing students to multiple-choice questions.

Fortunately, it is possible to reduce or eliminate the negative effects of multiple-choice questioning by providing feedback. Butler and Roediger (2008) and Butler et al. (2007) had students take initial multiple-choice tests and restudy correct answers as feedback. Both studies manipulated whether feedback occurred immediately, following each multiple-choice question, or after a delay, by having students answer all questions and then study the correct answers. On final short-answer tests, providing initial feedback cut the rate of lure intrusions in half (e.g., it was reduced from 20% to roughly 10% in Butler et al., 2007). Both immediate and delayed feedback formats were equally effective at reducing lure intrusions.

Turning now to the positive effects of multiple-choice questions, as noted earlier, answering multiple-choice questions tends to produce positive effects on subsequent retention. Researchers have assumed that multiple-choice questions might not be as beneficial as short-answer questions if multiple-choice questions require little retrieval effort—for instance, if students easily recognize correct answers and need not think much to retrieve prior knowledge. Therefore, recent work has examined ways to construct multiple-choice questions so that they afford greater retrieval effort.

Little et al. (2012) constructed multiple-choice questions that contained "competitive" incorrect alternatives, defined as alternatives that were plausible. For example, for the question "what is the hottest terrestrial planet?" (correct answer: Venus), Mercury and Mars represent plausible alternatives while Uranus and Saturn would be less plausible (Little and Bjork, 2015). Little et al. reasoned that when multiple-choice alternatives were plausible, subjects would need to retrieve the correct answer and also retrieve reasons why other alternatives were incorrect. Little et al. found that taking an initial multiple-choice test produced better performance on a final short-answer test than did taking an initial short-answer test. This pattern occurred when no feedback was provided following the initial test (Kang et al., 2007). More importantly, Little et al. showed that initial multiple-choice conditions outperformed initial short-answer conditions on final tests that included questions that were related but not directly tested on the initial test. This positive effect on related questions presumably occurs because of the effortful retrieval or competition afforded by plausible multiple-choice alternatives (see too Little and Bjork, 2015).

For practical purposes, multiple-choice questions produce positive effects on learning. There is a risk that students may pick up erroneous knowledge when they answer multiple-choice questions, but these negative effects are dramatically reduced when students receive feedback. Moreover, there may be unique benefits of multiple-choice questions when those questions lead learners to reason about why certain alternatives are incorrect.

2.27.4.5 Spaced Retrieval Practice

It is well known that spaced practice, in which time or other events occur between repetitions, promotes better learning and long-term retention than does massed practice, in which repetitions occur in immediate succession without intervening events. Spaced

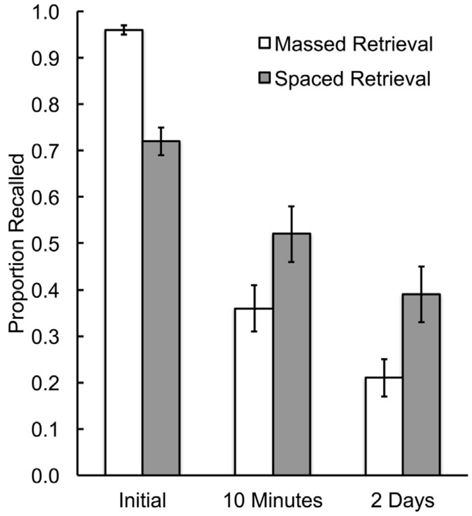

Figure 4 Proportion recalled in massed and spaced retrieval conditions during initial recall and on final recall tests after delays of 10 minutes or 2 days. Data from Experiment 3 in Karpicke, J. D., Roediger, H. L., 2007a. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. J. Exp. Psychol. Learn. Mem. Cogn. 33(4), 704–719.

retrieval practice refers to the advantage of distributing retrieval attempts over time relative to massing repeated retrieval attempts together. Like other manipulations discussed in this section, spaced retrieval offers less support and requires more effort relative to immediate, massed retrieval. However, as the lag between a study episode and an initial test increases, the likelihood of forgetting also increases. Thus, spaced retrieval is a prime example of how retrieval success and retrieval effort trade-off and how balancing the two is crucial for promoting learning.

An example of this trade-off in spaced retrieval comes from Karpicke and Roediger (2007a, Experiment 3). They had subjects learn vocabulary word pairs and varied the lag between when an item was studied and when it was retrieved. Their experiment included several conditions, but two are pertinent for present purposes: Some items were studied and then immediately recalled without any intervening items (a lag of 0), while others were studied and then recalled after five intervening items had occurred, introducing a small amount of spacing between initial study and retrieval. There was no feedback following retrieval trials. Fig. 4 shows the proportion of items recalled on the initial retrieval practice trials and on final delayed tests that occurred either 10 min or 2 days after the initial learning phase. Items were recalled nearly perfectly immediately after presentation, while inserting a small spacing produced a decrease in recall of approximately 25%. However, spaced retrieval enhanced retention on the final tests. The pattern of results clearly illustrates the trade-off of success and effort. Massed practice produced nearly perfect initial retrieval success but led to rapid forgetting as seen 10 min after the learning phase. Spaced retrieval practice, on the other hand, resulted in lower initial retrieval success but greater performance on final tests.

Is there a way to schedule spaced retrieval practice opportunities to create high levels of retrieval success and retrieval effort? Nearly 40 years ago, Landauer and Bjork (1978) proposed that an expanding pattern of retrieval practice might accomplish that. In an expanding spacing schedule, initial retrievals occur shortly after items are studied for the first time, and the lags between later repeated retrievals grow increasingly longer. For example, a subject might recall an item immediately after studying it (lag 0), then after one intervening item, then again after five more items, and again after nine more items. Thus, the spacing intervals expand between repeated retrieval attempts (0-1-5-9). The idea was that an expanding schedule would afford high levels of retrieval success, because early recalls occur shortly after study, while repeated retrievals grow increasingly more difficult because of increased spacing (for reviews, see Balota et al., 2007; Roediger and Karpicke, 2011).

The idea behind expanding retrieval is intuitive, and for years it was recommended as an optimal way to spaced repeated retrievals, even though there were hardly any studies that had tested the efficacy of the approach. However, expanding retrieval has received more examination over the past decade than it did in the decades following Landauer and Bjork's (1978) publication. Surprisingly, the bulk of recent work has suggested that expanding intervals may not represent the most effective way to schedule retrieval practice (see Balota et al., 2006; Carpenter and DeLosh, 2005; Cull, 2000; Karpicke and Bauernschmidt, 2011; Karpicke and Roediger, 2007a, 2010; Logan and Balota, 2008; Maddox et al., 2011; Pyc and Rawson, 2007; Storm et al., 2010). Expanding retrieval schedules (e.g., 0-1-5-9) are clearly superior to massed retrieval practice schedules (0-0-0-0), because while both conditions support high levels of retrieval success, the expanding schedule introduces spacing to promote learning. However, recent work has compared expanding schedules to other schedules that are matched on total spacing. For example, a 5-5-5-5 schedule, with five trials between each repetition, has the same total spacing as a 0-1-5-9 expanding schedule, but trials are equally spaced in the 5-5-5-5 schedule. The surprising finding in several recent studies has been that equally spaced schedules produce retention that is the same as or sometimes better than retention produced by expanding schedules (e.g., Karpicke and Roediger, 2007a; Pyc and Rawson, 2007).

A frequent problem in research on spaced retrieval is that expanding and equally spaced schedules afford different levels of initial retrieval success. Any advantage of an expanding schedule may occur because subjects recall more items in an expanding condition, rather than because the interval between repeated retrievals gradually increases. Researchers have attempted to address this methodological issue in various ways (see Carpenter and DeLosh, 2005; Karpicke and Roediger, 2007a). Karpicke and Bauernschmidt (2011) carried out one large experiment that examined several spaced retrieval schedules and minimized differences in initial retrieval success across spacing conditions. Students studied and recalled a set of vocabulary word pairs across alternating study and test trials and continued until they had recalled each item (following similar procedures by Karpicke, 2009; Karpicke and Roediger, 2008). When an item was recalled for the first time, it was assigned to one of several spaced retrieval conditions. This procedure ensured that students had successfully recalled items in all spaced retrieval conditions. Any differences across spacing conditions could be attributed to the spacing schedules, rather than to different levels of retrieval success across conditions. The effects of various spaced retrieval conditions were assessed on a final cued recall test 1 week after the initial learning session.

Fig. 5 shows the final test results from Karpicke and Bauernschmidt's (2011) experiment. There are several things to glean. The left portion of the figure shows three control conditions. In one condition, the subjects simply studied the items one time, which led to very poor retention. Another group recalled each item once but did not repeatedly recall, and a third group repeatedly recalled three times in a massed fashion. Practicing to the criterion of one correct recall of each item enhanced retention, relative to studying the items only one time. Massed repeated retrieval produced no benefit relative to dropping items after one recall. The remaining data in **Fig. 5** demonstrate the effects of spaced retrieval practice. Overall, increasing the absolute or total spacing across repeated trials enhanced long-term retention. However, there was no discernable difference between the three relative spacing conditions, expanding, equally spaced, or contracting.

Two recent studies are noteworthy because they examined spaced retrieval schedules over days, instead of manipulating the lags between items within a learning session, and assessed retention several weeks after the final learning session. Kang et al. (2014) had students learn vocabulary word pairs in an initial session followed by three repeated sessions. In each session, students completed three cycles of retrieval practice with feedback. The repeated sessions were spaced over several days according to an expanding schedule (2-6-19 days between sessions) or an equally spaced schedule (9-9-9 days). A final test occurred 56 days after the last practice session. Kang et al. found a 3% advantage of the expanding schedule relative to the equally spaced schedule, a difference that did not reach statistical significance.

Kupper-Tetzel et al. (2014) had subjects learn unrelated word pairs in an initial learning session and two relearning sessions. In each session, subjects completed retrieval practice with feedback trials and continued until they correctly recalled each item two times. The relearning sessions were either expanding (1 day, then 5 days), equally spaced (3-3 days), or contracting (5-1 days). Kupper-Tetzel et al. assessed different groups of subjects at one of several different retention intervals: 15 min, 1 day, 7 days, or 35 days after the last learning session. Overall, they found no evidence for the superiority of an expanding retrieval schedule. In fact, the contracting schedule produced the best performance at the 1-day and 7-day retention intervals (the expanding and equally spaced conditions were best at the 35-day retention interval, and performance did not differ in those two conditions).

In sum, spaced retrieval practice invariably enhances learning relative to massed practice. Expanding retrieval schedules, where the interval between successive retrieval trials gradually increases, would seem to be an intuitive way leverage the benefits of both retrieval success and retrieval effort. However, a wealth of research has failed to demonstrate consistent advantages of expanding

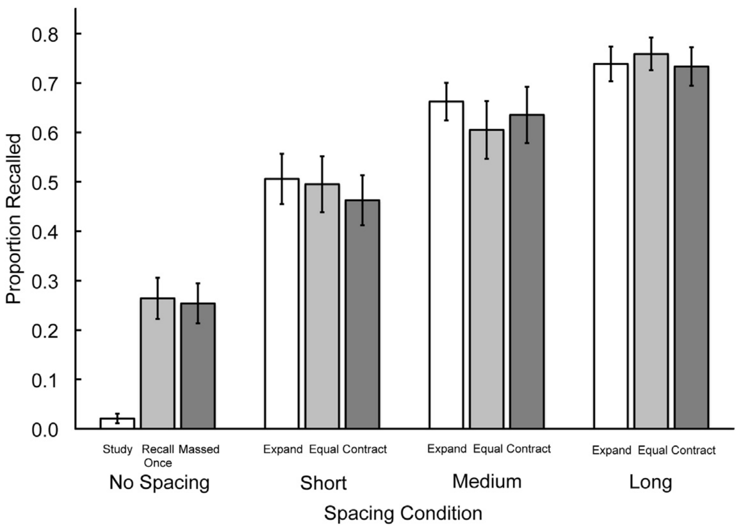

Figure 5 Proportion recalled on a week-delayed final test as a function of initial spaced retrieval practice with different absolute spacing conditions (short, medium, or long) and relative spacing conditions (expanding, equal, or contracting). Data from Karpicke, J.D., Bauernschmidt, A., 2011. Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. J. Exp. Psychol. Learn. Mem. Cogn. 37 (5), 1250–1257.

schedules of retrieval practice relative to equally spaced or contracting schedules. In the extant literature, there is not strong evidence that one schedule of spaced retrieval is superior to another.

2.27.4.6 Direct Manipulations of Initial Episodic Retrieval

In the literature reviewed so far, the challenge of balancing retrieval success and retrieval effort is evident in a wide range of retrieval practice scenarios. The conditions an experimenter wishes to compare may differ in initial retrieval success or reexposure to the materials (e.g., comparing restudy to recall conditions, or multiple-choice to short-answer conditions, or massed retrieval to spaced retrieval conditions). When there is such a difference, this clouds interpretation of performance on a final test. In the studies reviewed in the present section, researchers have attempted to hold everything constant, including initial retrieval success and reexposure to the materials, and manipulate whether people recollect a prior study context during retrieval practice. Although only a few studies have used this approach, these studies speak directly to the episodic context account of retrieval-based learning (Karpicke et al., 2014b), which proposes that reinstating a prior study context is crucial for retrieval practice effects.

One might imagine that comparing a restudy condition to an initial yes/no recognition condition would serve as a straightforward manipulation of whether subjects recollect the study context (Carpenter and DeLosh, 2006; Glover, 1989; Hogan and Kintsch, 1971; Mandler and Rabinowitz, 1981). There are at least two problems with this comparison. First, recognition tests do not necessarily require subjects to recollect a prior study context. People regularly rely on automatic retrieval processes, characterized as familiarity, fluency, or knowing, during recognition tests (Jacoby, 1991; Rajaram, 1993; Yonelinas, 2002). Second, standard yes/no recognition tests include lures, while restudy control conditions typically do not. For example, Carpenter and DeLosh (2006, Experiment 1) found that an initial yes/no recognition test was consistently worse than restudy on all final assessments (see **Table 1**). However, whereas subjects in their restudy condition restudied only the list of eight target items, subjects in the recognition condition were exposed to the eight targets and eight lures. The detrimental effect of a standard yes/no recognition test may have been due to the increased list length in the recognition condition.

Other work has attempted to hold everything constant across conditions (e.g., list length and reexposure to the target items) and directly manipulate whether people retrieved prior episodic or contextual details about the original study event. Karpicke and Zaromb (2010) accomplished this by having subjects generate targets or retrieve target words from a prior study list. In their experiments, subjects studied a list of target words and then either restudied the words, generated the words, or recalled the words. Specifically, subjects either restudied the targets paired with a related cue (e.g., for the target word *love*, subjects restudied *heart—love*) or were shown cues and fragments of the targets (e.g., *heart—l_v_*). In a generate condition, subjects were instructed to complete the fragment with the first word that came to mind that successfully completed it – a standard instruction used in studies of priming on implicit memory tests that does not require subjects to think back to a prior event. In the recall condition, subjects were told to think back to the original study phase and remember a word that completed the fragment – an instruction to explicitly remember the study episode. Finally, in the third phase of the experiments, subjects took a criterial free recall or recognition test. On the final tests, performance was consistently better when subjects had practiced retrieval (the recall condition) relative to when they generated the targets without thinking back to the original study episode (the generate condition). Importantly, performance in the second phase of the experiments was roughly the same in the generate and the recall conditions (about 75% correct), indicating that differences on the final criterial test were not due to differences in reexposure to items in the second phase. The only difference between conditions was whether subjects were asked to think back to original study context. Doing so enhanced subsequent retention, consistent with the episodic context account of retrieval-based learning (see also Pu and Tse, 2014).

In more recent work, Whiffen and Karpicke (2017) manipulated whether subjects recollected an initial study episode by having them make list discrimination judgments. In their experiments, subjects studied two short lists of words, separated by a distracter task, and then were shown the words again with the two lists mixed together. In a restudy condition, subjects were simply told to restudy the words. In the retrieval practice condition, subjects made a list discrimination judgment for each word by indicating whether the word occurred in the first or second list. Thus, subjects in both conditions reexperienced all items, but people who made list discrimination judgments were required to think back to the original study phase and retrieve details about when the word had occurred. On a final free recall test, the list discrimination condition produced about a 10% advantage relative to restudy. Whiffen and Karpicke also examined additional measures on the final test (see Ancillary Measures on Criterial Assessments), which showed that initial retrieval practice enhanced the degree to which subjects organized their recall according to the original episodic/temporal order of items.

At the moment, the literature contains only a few studies that have directly manipulated initial episodic retrieval while holding all other factors constant. Those studies have provided consistent evidence that inducing subjects to recollect the prior episodic context of an event enhances subsequent retention.

2.27.5 Characteristics of Final Assessments of Learning

In research on retrieval-based learning, characteristics of how learning is assessed on a final criterial assessment fall into three broad categories. First, several studies have examined how initial retrieval practice affects retention after a delay to establish that effects of retrieval practice are long lasting. Second, many studies have included ancillary measures during the final assessment, including measures of how learners organize their output and measures of response times. Third, a major emphasis has been on transfer of learning, where the questions or conditions on the final assessment differ from the exact conditions experienced during initial learning.

2.27.5.1 Retention Interval

A remarkable feature of retrieval-based learning is that the benefits of retrieval practice persist over long retention intervals. Many studies of retrieval practice use final assessments given several days or a week after initial learning. Some studies have demonstrated that the benefits of retrieval practice can be especially long lasting. For example, in research with medical students, Larsen et al. (2013) found that initial testing of clinical neurology topics enhanced retention 6 months later, relative to initially studying a review sheet. In research with middle-school students, Carpenter et al. (2009) found that initial retrieval practice of history facts enhanced retention relative to repeated study on a final test given 9 months after the initial learning session.

The benefits of retrieval practice appear to grow larger with longer retention intervals. In Rowland's (2014) meta-analysis, based on data from 159 effect sizes, retrieval practice effects were larger at retention intervals greater than 1 day ($g = 0.69$) relative to retention intervals less than 1 day ($g = 0.41$). If repeated retrieval alters the course of forgetting, as depicted in Hanawalt's (1937) data displayed in **Fig. 1**, then the gap between retrieval practice conditions and control conditions exhibiting normal forgetting should indeed grow larger at longer retention intervals.

However, the finding that retrieval practice effects grow larger at longer retention intervals has led some researchers to suggest that retrieval practice effects only occur after a delay, and that there are no benefits at short retention intervals. This conclusion is mistaken. Rowland's (2014) analysis indicates that there are reliable positive effects at short delays. It is true that some forgetting needs to set in, so that final test performance is not at ceiling, for retrieval practice effects to be observed in the short term. Conditions that ensure high levels of initial retrieval success, which is crucial for retrieval practice effects, are indeed likely to lead to ceiling effects in the short term. Nevertheless, several experiments have demonstrated retrieval practice effects on final assessments that occurred only a few minutes after the initial learning phase (Karpicke and Zaromb, 2010; Rowland and DeLosh, 2015; Smith et al., 2013; Verkoeijen et al., 2012).

The question of whether initial retrieval practice alters the rate of forgetting has been a matter of some debate. For instance, Kornell et al.'s (2011) bifurcation idea proposes that retrieval practice does not alter the forgetting rate for retrieved items. According to that account, in a retrieval practice condition, a few items (ones that are successfully recalled initially) receive a large degree of initial strengthening, whereas in a repeated study condition, all items that are restudied receive a smaller degree of strengthening. The account assumes that all items are forgotten at the same rate, but because retrieval practice items accrued more strength initially, they are recalled better than restudy items on delayed tests. The bifurcation account essentially considers the benefit of retrieval practice to be an artifact of item selection (see Slamecka and Katsaiti, 1988; Thompson et al., 1978).

Two lines of evidence suggest that retrieval practice does indeed alter the rate at which items are forgotten. First, several experiments have had subjects learn lists to criterion, for instance, by studying and recalling items until each item has been correctly recalled. Once items have been recalled, manipulations of repeated retrieval practice or repeated studying are introduced. In these experiments, all items are initially "learned" because each item is recalled once during the learning phase. Indeed, this was the traditional method researchers used to ensure appropriate study of forgetting rates (Underwood, 1964). Under these conditions, any effect of repeated retrieval must be attributed to changes to the forgetting rate for repeatedly retrieved items. These procedures tend to show rather sizable effects of repeated retrieval on long-term retention (Karpicke, 2009; Karpicke and Roediger, 2007b, 2008; Karpicke and Smith, 2012).

The second line of evidence comes from research that has examined mathematical properties of forgetting curves in repeated study and repeated retrieval conditions. Carpenter et al. (2008) carried out three experiments in which subjects either restudied or retrieved items and took a final test at one of several retention intervals (5 min, 1 day, 2, 7, 14, or 42 days), which allowed them to fit the pattern of forgetting over time to a power function. By using a quantitative approach to examine the mathematical form of forgetting, Carpenter et al. determined that initial retrieval practice did indeed alter the rate at which items were forgotten.

2.27.5.2 Ancillary Measures on Criterial Assessments

Several studies have examined additional measures during a final criterial test to gain theoretical leverage on the nature of retrieval-based learning. Four ancillary measures have received the most attention: (1) memory for mediators and with mediators as retrieval cues; (2) memory for details of the initial episodic or temporal context; (3) retrieval dynamics and response times; and (4) measures of semantic organization.

Such ancillary measures may prove interesting and informative, but the results must be interpreted with caution to avoid a beguiling reasoning error. The faulty logic is that if retrieval practice improves performance on an ancillary measure during a final test, then the construct being assessed with that measure must have caused the retrieval practice effect. This chain of reasoning amounts to affirming the consequent. For example, one might reason that if retrieval-based learning is due to the production of mediators, then final memory for mediators should be enhanced in a retrieval practice condition. If the latter is true—retrieval practice enhances final memory for mediators—that does not necessitate that retrieval-based learning was caused by the production of mediators. The ancillary measures reviewed in this section may help discriminate among theories of retrieval-based learning, but a degree of caution should remain in mind.

Several studies have examined memory for mediators on a final test. In experiments with word pair materials, mediators are words that are strongly associated to cues but not associated to targets (e.g., *board* is a mediator for the pair *chalk–crayon*). Carpenter (2011) found that people were more likely to falsely recognized mediators on a final recognition memory test following retrieval practice relative to restudy. Pyc and Rawson (2010) showed that there was a retrieval practice effect when mediators were used as cues on the final test (e.g., *board–?* as a cue to recall *crayon*). Additional research has also observed effects on mediator-cued final tests (Carpenter, 2011; Rawson et al., 2015a). Coppens et al. (2016), however, concluded that this effect may be smaller than what original reports concluded (the true effect size may be in the 0.10–0.20 range). These findings on final mediator-cued tests are consistent with the elaborative retrieval account of retrieval-based learning: If mediators come to mind during initial retrieval, then it is reasonable to expect that mediators would function as effective retrieval cues and might be falsely recognized. However, as noted earlier, these findings do not necessitate an explanation wherein mediators occur during initial retrieval practice (see Lehman and Karpicke, 2016).

Other studies have examined final test conditions aimed at assessing recollection of initial episodic or contextual details. These studies have reported that initial retrieval practice enhances a person's ability to remember details of the initial learning context. For example, Chan and McDermott (2007) had subjects study two word lists and either practice free recall after each list or complete a filler task. They then had subjects take a final list discrimination test, requiring subjects to indicate whether words had occurred in the first or second list. Initial retrieval practice enhanced performance on the criterial list discrimination test. Brewer et al. (2010) found similar results, showing that initial retrieval enhanced final list discrimination performance but did not improve memory for other source attributes (e.g., the gender of a speaker's voice; see also Rosburg et al., 2015). Verkoeijen et al. (2011) had subjects make remember/know judgments on a final test (Tulving, 1985) and found that initial retrieval practice enhanced the subjective experience of remembering on the final test. These findings have been interpreted as consistent with the episodic context account of retrieval-based learning (Karpicke et al., 2014b): If initial retrieval practice involves reinstating and updating context representations, then retrieval practice should enhance performance on final tests that assess recollection of episodic or temporal details. However, the criticisms applied earlier apply here as well. The finding that retrieval practice enhances final context memory does not necessitate a theoretical explanation whereby initial contextual retrieval produces the benefits of retrieval practice.

A related final assessment method deserves mention, whereas the assessments described above measured the ability to determine temporal order between lists (which list a word originally occurred in), one study has examined the effect of retrieval practice on within-list order memory. Karpicke and Zaromb (2010, Experiment 3) had subjects restudy, generate, or practice retrieving eight-item word lists. The subjects then completed a final order reconstruction test, in which subjects were shown the words in a random order and told to sort them into their original temporal order. It is well known that generating items disrupts within-list order memory relative to reading items (McDaniel and Bugg, 2008; Nairne et al., 1991). Karpicke and Zaromb found that retrieval practice also disrupted order reconstruction performance, just as generation did. At the moment, there is no detailed explanation for why retrieval practice might enhance between-list order memory but disrupt within-list order reconstruction.

A few studies have examined the effects of retrieval practice on retrieval dynamics or response times during a final test. The episodic context account proposes that updated context representations, produced via repeated retrieval practice, facilitate memory search processes during subsequent retrieval. More efficient memory searches should be evident in patterns of response times and analyses of retrieval dynamics. Relatively few studies have reported response times during final tests, but the existing literature indicates that response times during final tests are faster in retrieval practice conditions than they are in restudy control conditions (Keresztes et al., 2014; van den Broek et al., 2014; van den Broek et al., 2013). Lehman et al. (2014) assessed retrieval dynamics on a final test with cumulative recall curves, which measure free recall output over time (Wixted and Rohrer, 1993). They found that initial retrieval practice led to rapid recall of several items early on during final free recall, relative to conditions where subjects only studied items or completed an elaborative study task. This pattern of retrieval dynamics suggests that retrieval practice produced a well-defined memory search set that facilitated retrieval on the final test. Whiffen and Karpicke (2017) examined the effect of initial retrieval practice on several aspects of final recall. They found that retrieval practice enhanced the degree to which recall was clustered around the original temporal organization of items (see Sederberg et al., 2010) and the degree to which people used temporally defined search strategies during recall (based on an information foraging analysis; see Hills et al., 2015).

Finally, several studies have assessed the effects of initial retrieval practice on semantic organization during a final free recall test. Typically, these studies have used categorized word lists and calculated clustering scores on the final test (Roenker et al., 1971). Across a handful of experiments, the pattern of results is mixed, and experimental outcomes may hinge upon how initial retrieval practice is implemented. Zaromb and Roediger (2010) and Congleton and Rajaram (2012) found that initial retrieval practice enhanced final semantic organization. In those studies, subjects practiced retrieval in initial free recall tasks. Peterson and Mulligan (2013, Experiment 2) and Mulligan and Peterson (2015) obtained contrary results, finding that initial retrieval practice produced lower semantic clustering scores relative to restudy conditions. However, these authors also reported that initial retrieval practice produced worse final recall performance relative to initial restudy. The Peterson and Mulligan experiments used word pair materials, sometimes with rhyming pairs (e.g., *moon–spoon*, *wife–knife*), and implemented retrieval practice with initial cued recall tests. Whiffen and Karpicke (2017) found that initial retrieval practice with a list discrimination task produced lower scores on measures of semantic organization during final recall, relative to restudy conditions. It may be that initial free recall promotes subsequent semantic clustering while initial item-specific recall tasks, like cued recall and list discrimination, do not. However, other experiments have complicated this simple conclusion. For example, Knouse et al. (2016) and Lipowski et al. (2014) used initial free recall retrieval practice conditions but did not observe benefits on final semantic clustering scores. Furthermore, in a series of experiments, Rawson et al. (2015b) failed to replicate the "negative testing effects" observed by Peterson and Mulligan, finding instead that cued recall retrieval practice produced modest enhancements to final category clustering. Looking across the literature, it is unclear whether initial retrieval practice enhances final assessments of semantic organization in free recall.

To summarize this fairly brisk tour through a wide range of ancillary measures: Retrieval practice produces consistent benefits on final tests that are sensitive to recollection of contextual details, and it enhances the efficiency of final memory searches, as assessed with measures of response times and retrieval dynamics. The effects of retrieval practice on final memory for mediators and on mediator-cued final tests have been positive, though the effects may not be large (Coppens et al., 2016). The effects of retrieval on final measures of semantic organization are decidedly unclear. Such effects may depend on the way initial retrieval is carried out, but further work to clarify matters is in order.

2.27.5.3 Transfer of Learning

A central focus in recent research on retrieval practice has been to establish the benefits of retrieval-based learning on measures of "meaningful learning." Broadly, meaningful learning is thought to have occurred when students possess organized, coherent, and integrated mental models that allow them to make inferences and apply their knowledge (Mayer, 2008). With this outcome in mind, several studies have examined *transfer*, which refers to situations where initial retrieval practice conditions differ from conditions during a criterial assessment of learning. In one sense, all learning and memory performance is transfer, because the past never repeats itself and learners must always use the cues available in a current context to retrieve and reconstruct prior knowledge and experiences (Karpicke, 2012; McDaniel, 2007). Carpenter (2012) reviewed the literature on the effects of retrieval practice on transfer and identified three ways retrieval practice has been shown to promote transfer: transfer across "temporal contexts," in reference to the fact that retrieval practice promotes long-term retention (see the earlier section on Retention Interval); transfer across test formats; and transfer across knowledge domains. The present section is focused on the latter two approaches. Research on retrieval-based learning has examined transfer in two general ways: One approach has been to examine final tests that include inference and problem-solving questions, and another approach has examined whether practicing retrieval of one portion of material transfers and aids retention of other, nontested portions of material.

One straightforward way to examine transfer is to vary whether the format of initial test questions exactly matches the format of final test questions. Research comparing short-answer and multiple-choice questions, reviewed earlier in this chapter (see Retrieval Practice With Initial Short-Answer and Multiple-Choice Formats), has found that matching initial and final test formats does not matter much for retrieval practice effects. Other work has tackled the question of test format matching in other ways. Carpenter et al. (2006) had subjects practice retrieval of word pairs (e.g., after studying the pair *train–plane*, subjects were given the cue *train* to retrieve the target *plane*). They found that practicing retrieval enhanced performance on a final test regardless of whether the cue word was given to recall the target (*train–?*) or the target was given to recall the cue (*plane–?*). Similarly, McDaniel et al. (2007a) used fill-in-the-blank questions and varied the exact wording of initial and final test questions by omitting different words on the two tests. They observed retrieval practice benefits even though the test formats did not exactly match (see also McDaniel et al., 2013). Rohrer et al. (2010) had students learn the names and locations of cities on a map. The students practiced retrieval by matching the city names to locations on the maps, and this retrieval practice activity enhanced performance on a transfer test where students recalled spatial locations of cities that lay along routes connecting other cities. Finally, several other studies have shown that practicing initial free recall of texts enhances performance on final short-answer questions (e.g., Blunt and Karpicke, 2014; Karpicke and Blunt, 2011; Smith et al., 2016). In sum, the benefits of retrieval practice do not depend on an exact match between initial retrieval practice conditions and the final criterial test format.

Several studies have assessed transfer by varying the types of questions students must answer on criterial assessments. An example of the benefits of retrieval practice for transfer comes from a series of experiments by Butler (2010). In one experiment, students studied educational texts and either reread the texts three times or took three initial short-answer tests (with feedback) in retrieval practice conditions. Butler examined two initial retrieval conditions: In a same-test condition, students answered questions that were phrased identically on each test, whereas in a variable-test condition, students answered questions that assessed the same concepts but were rephrased on each test. One week later, students took final tests that included two types of questions: factual questions that pertained to only one fact from the materials, and conceptual questions that pertained to concepts that learners had to abstract from multiple portions of the texts. Importantly, all of the questions on the final test were new inferential questions, because the questions required learners to make inferences that went beyond what was stated directly in the materials. **Fig. 6** shows performance on the final test. Practicing retrieval enhanced performance on the final inference questions, relative to restudying. Interestingly, variable initial test conditions did not lead to better performance than same-test conditions.

Several additional studies have examined the effect of initial retrieval practice on performance on final problem-solving or inference questions. Many of these studies have used free recall as the initial retrieval practice activity (Roediger and Karpicke, 2006b). McDaniel et al. (2009) showed benefits of a single initial retrieval—freely recalling material in a "read–recite–review" procedure—on final problem-solving questions (Mayer and Gallini, 1990). Karpicke and Blunt (2011) showed benefits of initial free recall retrieval practice on final inference questions and final assessments that required learners to create concept maps as the assessment (for additional studies with final inference questions, see Blunt and Karpicke, 2014; Johnson and Mayer, 2009; Smith et al., 2016; Smith and Karpicke, 2014).

Hinze and Wiley (2011, Experiment 3) had students read educational texts and compared initial fill-in-the-blank tests to paragraph recall, where students wrote down as much of material as they could. Importantly, Hinze and Wiley examined retention 2 days after initial learning with a final multiple-choice test that included transfer questions, where the wording of the final test

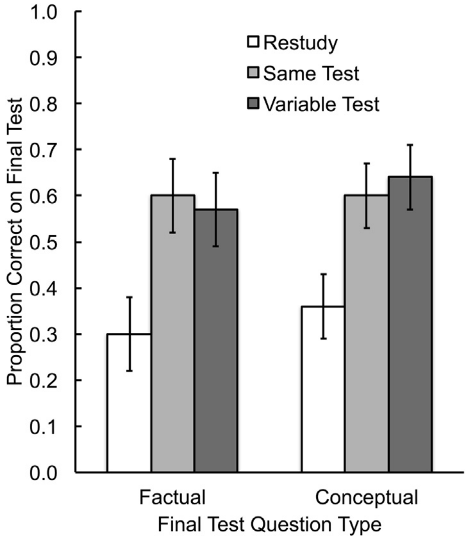

Figure 6 Proportion correct on a final transfer test, containing factual and conceptual questions, for students who repeatedly studied or took initial tests. Data from Experiment 1 in Butler, A. C., 2010. Repeated testing produces superior transfer of learning relative to repeated studying. J. Exp. Psychol. Learn. Mem. Cogn. 36(5), 1118–1133.

question differed from how concepts were described in the initial materials. The key finding was that the paragraph recall condition outperformed the fill-in-the-blank condition on the final transfer test, demonstrating another advantage of initial retrieval practice with relatively less support. In subsequent research, Hinze et al. (2013) found that instructing students to construct explanations during initial retrieval produced an additional benefit beyond initial free recall, as assessed on a final test that contained inference questions. However, Smith et al. (2016) found that prompting students to describe and explain concepts did not confer benefit beyond initial free recall. Free recall is an effective retrieval practice strategy for promoting long-term transfer to inference and problem-solving questions, but further research is needed on possible ways to extend the effectiveness of free recall.

In most of the studies cited earlier, the final inference questions required students to make deductions: Several premises were stated in the materials that students practiced retrieving, and students had to draw conclusions based on the statements from the original text to answer the final inference questions. It was therefore surprising, given the context of prior research, when Tran et al. (2015) claimed that retrieval practice did not enhance the learning of deductive inferences. They had subjects study a series of sentences, one at a time, and practice retrieval on an initial fill-in-the-blank test that required subjects to recall a single word in each sentence. This retrieval practice task did not enhance performance on final inference questions, relative to a restudy control condition. However, Eglington and Kang (2016) noted that Tran et al.'s procedure probably did not encourage subjects to form relations across sentences in the material, whereas other research on this topic has done so (see Karpicke and Aue, 2015). Eglington and Kang used Tran et al.'s materials and found positive effects of retrieval practice on final inference performance when subjects studied the sentences simultaneously and when they practiced free recall.

Finally, researchers have examined whether practicing retrieval of one portion of material transfers and aids retention of other, nontested portions of material. Under certain circumstances, recalling only a portion of a set of items can block or induce forgetting of other items within the set (see Roediger and Neely, 1982). Bauml and Kliegl (this volume) provide an extensive review of this research on interfering effects of retrieval. Such interfering effects raise the possibility that practicing retrieval of a portion of material may not benefit nontested portions, or might hinder their learning. However, research on this topic has shown that the benefits of retrieval practice spread to portions of material that are not explicitly tested and has identified conditions under which these facilitation effects occur.

Chan et al. (2006) initiated research on the facilitatory effects of retrieval practice for nontested material. In a series of experiments, students read educational texts and took an initial short-answer test with questions that were designed to cover some portions of the material but not others. On a later criterial test, practicing retrieval enhanced performance on both the tested and nontested material, relative to control conditions that did not practice retrieval. Chan et al. called this phenomenon *retrieval-induced facilitation* (see also Little et al., 2011). Importantly, Chan (2010) showed that these transfer effects persist after delays of 1 day and 1 week. Subsequent work has also identified the importance of relational or integrative encoding of the materials for obtaining retrieval-induced facilitation. Chan (2009) demonstrated the importance of relational encoding by having students study texts where the order of sentence had been randomized, which disrupted relational encoding during study. For these texts, retrieval practice enhanced subsequent retention for tested portions of the material, but there was no longer a retrieval-induced facilitation effect for nontested material.

Chan et al.'s (2006) theory to explain retrieval-induced facilitation was that during the process of retrieval, people actively search for information that is related to the desired target information. This "broad search" during initial retrieval is assumed to spread and enhance the learning of portions of material that are not explicitly tested. In retrospect, Chan et al.'s theory is a clear precursor to the elaborative retrieval account of retrieval practice (Carpenter, 2009, 2011). Recently, Rowland and DeLosh (2014), based on a series of experiments with word-list materials, proposed that the temporal cooccurrence of information within an episodic context was crucial for retrieval-induced facilitation. There is a vast literature exploring the conditions under which retrieval impairs or improves subsequent retrieval of related material, and it is reviewed extensively in the chapter by Bauml and Kliegl (this volume). Suffice it to say that the phenomenon of retrieval-induced facilitation with educational texts represents another example of how retrieval practice can promote long-term transfer.

2.27.6 Generalizing Across Learner Characteristics, Materials, and Educational Contexts

A great deal of research over the past decade has been aimed at exploring the generality of retrieval practice effects. Specifically, it is important to demonstrate that retrieval practice works for a variety of learners, across a range of materials, and in authentic educational contexts. This section reviews progress in these areas.

2.27.6.1 Learner Characteristics

The majority of research on retrieval practice continues to employ college students as experimental subjects. However, recent work has now established that the benefits of retrieval practice generalize beyond college students, specifically to children, older adults, and patient populations. Additional work has examined whether retrieval practice generalizes across individual difference factors (e.g., executive function and reading comprehension).

Several recent studies have now established that retrieval practice works well with children. In research on retrieval practice with children, the materials and retrieval practice methods are quite heterogeneous (Aslan and Bauml, 2016; Bouwmeester and Verkoeijen, 2011; Fritz et al., 2007; Goossens et al., 2014a; Jaeger et al., 2015; Karpicke et al., 2014a; Lipko-Speed et al., 2014; Lipowski et al., 2014; Marsh et al., 2012). Furthermore, most experiments on retrieval practice with children have included feedback following initial retrieval practice, prohibiting any conclusions about whether children exhibit the same direct mnemonic benefits of retrieval that adults do. One exception is Karpicke et al. (2016), who had children, aged 9–11 years, restudy or practice retrieval of word pairs, without feedback, and found benefits of retrieval practice on final free recall and item recognition tests. A few studies have suggested that the benefits of retrieval become more evident with older children (e.g., age 8 years) than with younger children (e.g., age 6 years) (Aslan and Bauml, 2016; Lipowski et al., 2014). Nevertheless, the work has established that retrieval practice benefits are observed in children.

Likewise, several studies have established positive effects of retrieval practice in healthy older adult populations. For example, Meyer and Logan (2013) tested two younger adult samples (aged 18–25 years) recruited from university or community settings, respectively, and a sample of community dwelling adults aged 55–65 years. The subjects completed a retrieval practice procedure modeled on Roediger and Marsh (2005). Subjects read educational texts, then answered initial multiple-choice questions or restudied the texts, and took a final short-answer test (the same questions but without multiple-choice alternatives) after a delay of 5 min or 2 days. The results are shown in **Fig. 7**, which depict consistent benefits of retrieval practice over restudying in younger and older adult populations at both retention intervals. Several additional studies have shown benefits of retrieval practice with older adults in their late 60s and 70s (Balota et al., 2006; Bishara and Jacoby, 2008; Coane, 2013; Henkel, 2007, 2008; Logan and Balota, 2008; Maddox et al., 2011; Rogalski et al., 2014; Tse et al., 2010).

Given the robust effects of retrieval practice on learning, it is no surprise that investigators have wanted to examine the possibility that retrieval practice might improve learning in memory-impaired populations. Several recent studies have shown promising results using retrieval practice as a memory remediation technique. This work has been done with patients with multiple sclerosis (Sumowski et al., 2010a; Sumowski et al., 2013), traumatic brain injuries (Pastotter et al., 2013; Sumowski et al., 2010b), and dementia of the Alzheimer's type (Balota et al., 2006). Recent work has also shown benefits of retrieval practice in patients with language impairments, such as aphasia (Friedman et al., 2017; Middleton et al., 2015; Middleton et al., 2016).

The benefits of retrieval practice have been established across age ranges, from early childhood to older adulthood, and in memory-impaired populations. There has been less work, however, focused on identifying individual difference factors that might mediate the benefits of retrieval practice. One complication that arises in such research is that initial performance (i.e., the level of success during initial retrieval practice) may differ across different groups. For example, a high ability group might recall more during initial retrieval practice than a low ability group would. Any difference in the size of a retrieval practice effect across ability groups would be influenced by differences in initial retrieval success, making it difficult to interpret the final test results (see the earlier section on Balancing Retrieval Success and Retrieval Effort). Nevertheless, two individual difference factors have received attention in recent research: trait anxiety and working memory capacity.

A few studies have asked whether anxiety during a test situation might affect how much learning occurs during the test. Tse and Pu (2012) and Mok and Chan (2016) measured trait anxiety with a standardized assessment tool. Both studies found benefits of retrieval practice regardless of whether people scored lower or higher on the anxiety assessment. However, people who scored higher on the trait anxiety assessment tended to show smaller testing effects. In related work, Hinze and Rapp (2014) induced performance anxiety (Beilock et al., 2004) within a testing effect experiment and found that inducing test anxiety reduced the size of the testing effect. However, in all three studies examining the relation between anxiety and retrieval practice, high-anxiety students recalled less

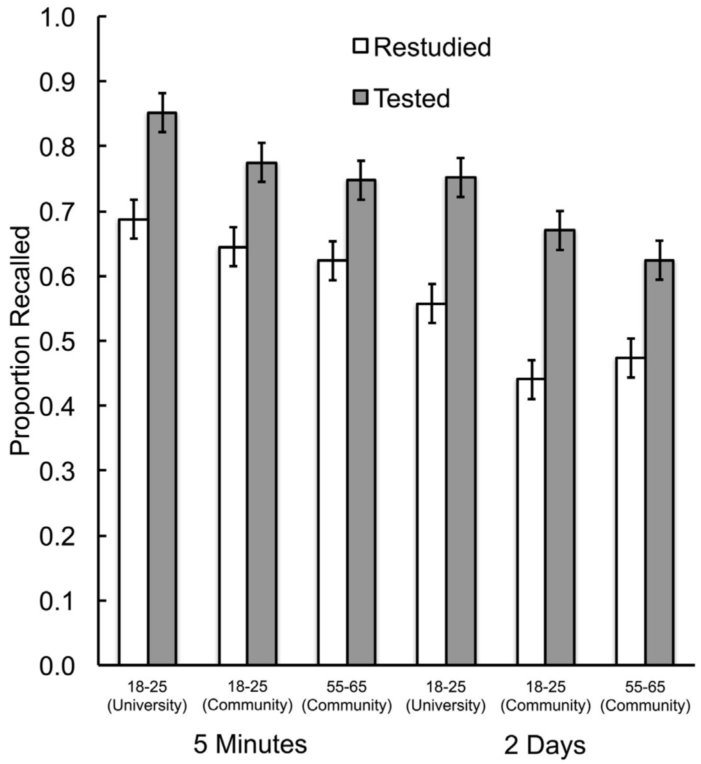

Figure 7 Proportion recalled on final tests in initial test vs. restudy conditions for younger adults (ages 18-25) and older adults (ages 55-65). Data from Meyer, A.N.D., Logan, J.M., 2013. Taking the testing effect beyond the college freshman: Benefits for lifelong learning. Psychol. Aging 28(1), 142–147.

during initial retrieval practice than low-anxiety students did. Thus, while it is clear that anxiety influences performance in retrieval practice experiments, it is not clear whether that is because it reduces initial recall or because it attenuates the mnemonic effect of retrieval on learning.

Finally, a handful of recent studies have assessed individual differences in working memory, or related measures of executive function (see Unsworth and Redick, this volume), to see whether high and low working memory subjects benefit differently from retrieval practice. At the moment, the literature has produced a mixed set of results. Some studies have found that low working memory subjects benefit more from retrieval practice than high working memory subjects do (Agarwal et al., 2016). Other studies have found the opposite pattern, where high working memory subjects benefit more than low working memory subjects do (Tse and Pu, 2012). Still other studies have found no relationship between working memory and retrieval practice (Brewer and Unsworth, 2012; Chan, 2009; Pan et al., 2015; Wiklund-Hornqvist et al., 2014). Brewer and Unsworth (2012) found that measures of working memory and attentional control were not related to the benefits of retrieval practice, but a separate measure of episodic memory ability was related, with low episodic memory subjects exhibiting larger retrieval practice effects. However, Pan et al. (2015) used the same measures and found no relationship between working memory and retrieval practice.

A persistent issue in research on individual differences in retrieval practice is that high and low ability subjects (e.g., high vs. low working memory subjects) tend to differ in the amount of material they recall during initial retrieval practice tasks. Differences in initial retrieval success across conditions clouds interpretation of final test performance (see Balancing Retrieval Success and Retrieval Effort for elaboration). Given the disparate set of findings in the existing literature, more work is warranted to clarify the relationship between working memory or executive function and retrieval-based learning.

2.27.6.2 Materials

Research over the past decade has successfully generalized the benefits of retrieval practice across a range of materials. Several examples provided so far in this chapter have reflected the array of materials used in retrieval practice research. Most research continues to use verbal materials—lists of words, word pairs, vocabulary words, facts, key term definitions, and educational texts. Retrieval practice has also been studied in experiments with visual materials like lists of pictures (Wheeler and Roediger, 1992), pairs of faces and names (Tse et al., 2010), and educational materials such as videos and multimedia presentations (Butler and Roediger, 2007; Johnson and Mayer, 2009). This section describes research with nonverbal materials and then provides a brief look at research with educational texts.

A few studies have examined retrieval practice with materials, such as abstract symbols or visual images, that are presumably not verbalizable. Kang (2010) had subjects practice retrieval of Chinese characters by mentally visualizing the characters during retrieval trials. He observed advantages of retrieval practice relative to a restudy condition with these materials. Coppens et al. (2011) used Adinkra symbols—visual symbols originally created by the Ashanti in Ghana—and found benefits of retrieval practice, with a procedure similar to Kang's. These findings of retrieval practice effects with materials that are not verbalizable (or that would be difficult to verbalize) pose a challenge to the elaborative retrieval account of retrieval-based learning (Carpenter, 2009, 2011). The elaborative retrieval account proposes that semantic elaboration afforded by retrieval cues is the mechanism responsible for retrieval practice effects. Because nonverbal materials afford few (if any) semantic elaborations, it is difficult to see how the generation of semantic elaborations or mediators would occur and create retrieval practice effects for these materials.

Other work has established that retrieval practice promotes learning of spatial materials. Rohrer et al. (2010), described earlier, showed retrieval practice effects in learning the locations of regions on maps. Carpenter and Pashler (2007) had students study maps that contained several landmarks such as roads, rivers, and trees. Following initial encoding of the maps, students who experienced maps with missing landmarks—thereby requiring the students to retrieve the landmarks—showed better retention when they were asked to redraw the maps on a final test relative to a restudy control condition. Carpenter and Kelly (2012) and Kelly et al. (2015) have examined retrieval practice within the context of three dimensional virtual reality displays, where subjects must learn the positions of objects in space (Carpenter and Kelly, 2012) or learn to navigate through a virtual building (Kelly et al., 2015). Both studies showed positive effects of retrieval practice on these spatial learning tasks.

Returning to research with verbal materials, considerable work during the past decade has focused on generalizing the benefits of retrieval practice to educational texts. The literature is now replete with examples in which retrieval practice enhances the learning of educational texts (e.g., McDaniel et al., 2009; Roediger and Karpicke, 2006b; Wissman et al., 2011). Little of this research, however, has manipulated attributes of texts to examine whether retrieval practice effects depend on text features. One exception is Karpicke and Blunt (2011, Experiment 2). They used texts with enumeration structures, which listed several facts and concepts about a topic, and sequential structures, which described sequences of events in a process. Karpicke and Blunt found equivalent retrieval practice effects for texts with different structures (see also Blunt and Karpicke, 2014).

Recently, a debate emerged regarding whether retrieval practice effects occur with complex materials, specifically materials that are high in "element interactivity" (van Gog and Sweller, 2015). Materials that are high in element interactivity contain several related ideas, so that learning of some ideas depends on learning other ideas in the material. Materials that are low in element interactivity contain ideas that can be learned in isolation, without reference to other ideas in the materials. van Gog and Sweller (2015) argued that retrieval practice effects do not occur with materials that are high in element interactivity, citing research by Tran et al. (2015) (discussed in section Transfer of Learning), research with more versus less organized texts (de Jonge et al., 2015), and research with worked example materials (Leahy et al., 2015; van Gog and Kester, 2012; van Gog et al., 2015).

The idea that retrieval practice effects may depend on the structure or complexity of materials is important and deserves further consideration. However, the research base cited by van Gog and Sweller (2015) is not convincing or definitive, as noted by Karpicke and Aue (2015) and Rawson (2015). First, two studies have examined retrieval practice with texts where the sentences of the texts are shown in correct order or in a random order. de Jonge et al. (2015) observed retrieval practice effects with a randomly ordered text but not with an intact text (a format that van Gog and Sweller argued was higher in element interactivity). Chan (2009) also manipulated whether texts were presented intact or with randomly ordered sentences and found robust retrieval practice effects for both formats. With only two studies examining this question, no firm conclusions can be made about whether retrieval practice effects differ for random versus intact texts. Notably, de Jonge et al.'s lack of a retrieval practice effect with an intact text does not fit with the wealth of research demonstrating retrieval practice effects with texts. Second, none of the three studies with worked example materials (Leahy et al., 2015; van Gog and Kester, 2012; van Gog et al., 2015) manipulated the complexity of materials. It is therefore possible that the retrieval practice conditions used in those studies may not have enhanced learning for any materials, even those that are easier or lower in complexity relative to worked example materials. At the moment, there is a wealth of evidence that retrieval practice enhances the learning of complex educational texts (Karpicke and Aue, 2015; Rawson, 2015).

In sum, retrieval practice effects generalize broadly across a wide range of materials. Most research on retrieval practice has aimed at generalizing to various material formats, rather than manipulating features of the materials and examining possible interactions (e.g., varying the relational structure of educational texts, as done by Chan, 2009; Karpicke and Blunt, 2011; de Jonge et al., 2015). Direct comparisons of retrieval practice effects with different materials are infrequent but could provide theoretical insights into mechanisms of retrieval-based learning.

2.27.6.3 Educational Contexts

The past decade has witnessed considerable effort toward generalizing retrieval practice to educational contexts. The general aim of such research has been to implement retrieval practice in actual courses with feasible instructional activities. The majority of this research has used classroom quizzes to implement retrieval practice, and this section reviews recent work on this topic (for a review of earlier research on classroom quizzing, see Bangert-Drowns et al., 1991).

A new method for implementing classroom quizzing or questioning has emerged during the past decade: student response systems, also known as clicker systems (Caldwell, 2007; Martyn, 2007; Mayer et al., 2009). Clicker systems allow instructors to pose questions (typically in multiple-choice format), record responses from all students regardless of class size, and provide immediate feedback, e.g., by displaying the correct answer on a screen in front of a classroom. Several studies in the past decade have used clicker systems to implement retrieval practice and have demonstrated their effectiveness in classrooms (Anderson et al., 2013; Campbell and Mayer, 2009; Mayer et al., 2009; McDaniel et al., 2011; McDaniel et al., 2013; Roediger et al., 2011). For example, McDaniel et al. (2011) used clicker systems to implement quizzes in eighth grade science classrooms for a range of different science topics (e.g., genetics, evolution, and anatomy). Students took multiple-choice quizzes using the clicker systems three times: immediately prior to lectures on the topics, immediately after the lectures, and 1 day prior to the exams covering the units. McDaniel et al. compared items from within each unit that were quizzed to items that were not quizzed, and the effects of quizzing were assessed on classroom exams that occurred about 20 days after material was introduced in class, at the end of the semester, and at the end of the school year. On all exams—even at the end of the school year—there were positive effects of initial quizzing, demonstrating long-lasting benefits of classroom clicker systems.

Other studies have used specialized computer software to implement classroom quizzing or other retrieval practice activities. Pennebaker et al. (2013) used a custom computer program to administer daily quizzes to students in large introductory psychology courses (with more than 900 students in the courses). Lindsey et al. (2014) created a personalized review system that implemented spaced retrieval practice tailored to each student's performance level. Lindsey et al.'s quizzing system enhanced student performance on end-of-semester exam in a middle school foreign language course. Butler et al. (2014) created a computer-based learning system that implemented repeated spaced retrieval practice with feedback and enhanced student learning in a college engineering course on signal processing.

However, clicker systems and specialized software are not necessary to implement retrieval practice in classrooms. Many researchers have implemented daily or weekly quizzing, using only paper-and-pencil materials, and observed positive effects on student learning. For example, Lyle and Crawford (2011) found that daily quizzes at the end of each lecture in a college statistics class improved students' performance on classroom exams. Similarly, Batsell et al. (2016) had students in one section of introductory psychology course take daily quizzes while students in another section of the course, taught simultaneously, did not. Students who took daily quizzes performed better on classroom exams that included questions that were identical to the quizzed questions, similar to quizzed questions (i.e., questions that covered topics assessed on the quiz), and completely new questions that were not directly related to topics covered on quizzes. Thus, daily quizzing had benefits that extended beyond mere repetition of quizzed material.

The benefits of frequent classroom quizzing have been seen in middle school, high school, college classrooms, and across a range of content areas. **Table 2** presents a compendium of recent studies showing benefits of classroom quizzing across grade levels, content areas, and quizzing method (paper, clicker, or computer-based quizzing). This spate of research over the past decade has firmly established that classroom quizzing is an effective way to leverage retrieval practice and improve student learning.

Finally, it is reasonable to wonder about possible negative effects of increased classroom quizzing. Specifically, for students who experience test anxiety, frequent classroom quizzing may exacerbate the anxiety these students experience in school. The existing research (reviewed in the earlier section on Learner Characteristics) suggests that while students who experience test anxiety benefit from retrieval practice, they may not benefit as much as students who do not experience anxiety. Fortunately, rather than increasing anxiety, frequent classroom quizzing appears to produce the opposite effect of reducing students' test anxiety (Leeming, 2002). In perhaps the most extensive study of this effect, Agarwal et al. (2014) surveyed 1408 middle and high school students who had experienced classroom-based retrieval practice programs (e.g., like the one implemented by McDaniel et al., 2011). The key findings from their survey were that 92% of students viewed the classroom retrieval practice activities positively, believing that retrieval practice helped them learn, and 72% said that frequent retrieval practice helped them feel less nervous about classroom exams. In addition to providing direct benefits to student learning, frequent quizzing also appears to benefit students by reducing test anxiety.

Table 2 Examples of recent studies showing benefits of classroom quizzing

Publication	*Grade*	*Course content*	*Quiz method*
Batsell et al. (2016)	College	Introductory psychology	Paper
Butler et al. (2014)	College	Engineering	Computerized quizzing
Carpenter et al. (2016)	College	Introductory biology	Paper
Hopkins et al. (2016)	College	Precalculus	Paper
Jensen et al. (2014)	College	Introductory biology	Computerized quizzing
Leeming (2002)	College	Introductory psychology	Paper
Lindsey et al. (2014)	Middle school	Foreign language	Computerized quizzing
Lyle and Crawford (2011)	College	Statistics	Paper
Mayer et al. (2009)	College	Educational psychology	Clickers
McDaniel et al. (2011)	Middle school	Science	Clickers
McDaniel et al. (2013)	Middle school	Science	Clickers
McDermott et al. (2014)	Middle school	Science	Clickers
Pennebaker (2013)	College	Introductory psychology	Computerized quizzing
Roediger et al. (2011)	Middle school	Science	Clickers

2.27.7 Conclusions and Future Directions

Roughly a decade ago, research on retrieval-based learning gained renewed interest among researchers interested in applying cognitive psychology to education (see Roediger and Karpicke, 2017). Roediger and Karpicke (2006a) provided an historical review of what was known at the time, based on research dating back to the early 20th century. The present chapter reviewed a great deal of the theoretical and empirical progress that has occurred over the past decade. In a decade of research, new methodological rigor has been applied to the study of retrieval practice. New theories of retrieval-based learning have been introduced, developed, and tested. Several attributes of initial retrieval practice activities have been examined. The benefits of practicing retrieval have been replicated hundreds of times across a range of initial retrieval tasks, and the effects of retrieval are long lasting and promote transfer, inferencing, and knowledge application. Retrieval practice has been shown to enhance the learning of a wide variety of materials for learners of nearly all ages and ability levels, and the effects have been established in classroom settings with authentic course content.

Despite this tremendous effort and progress, there is considerable room for additional progress in addressing several puzzles, inconsistencies, and gaps in the literature, many of which were noted throughout the chapter. There are many avenues for future research on retrieval-based learning, but this chapter concludes by highlighting four possible avenues.

First, there is still a need to deepen understanding of the mechanisms of retrieval-based learning. New theories that delineate possible mechanisms have been proposed during the past decade, but the theories are quite new and require additional scrutiny. Furthermore, it is essential to link theories of retrieval-based learning to global theories and models of learning, memory, and cognition.

Second, several puzzles persist about how to strike the appropriate balance between retrieval success and retrieval effort. For instance, while short-answer tests were thought to promote more learning than multiple-choice tests and were recommended a decade ago (McDaniel et al., 2007b), further research has called this conclusion into question. Likewise, expanding schedules of retrieval practice would appear to create an ideal balance of retrieval success and retrieval effort, but the bulk of research has not found expanding retrieval to be a superior form of spaced retrieval. The optimal balance between retrieval success and retrieval effort is not entirely clear-cut.

Third, the generality of retrieval practice is now well established across materials, final assessments, and (to a lesser extent) learner characteristics. Additional progress would be gained by examining retrieval practice from a contextual perspective on learning strategies (McDaniel and Butler, 2011; McDaniel and Einstein, 1989). Rather than seeking to generalize retrieval practice across learners, materials, and assessments, a contextual perspective would seek to identify whether retrieval practice effects are invariant or interact with these factors, for instance, by manipulating features of materials or aspects of final assessments. Some work discussed in this chapter has been done from this perspective, but it has not been a major focus of retrieval-based learning research.

Finally, there remains a pressing need to integrate retrieval practice into existing educational activities, or to develop new activities that incorporate retrieval practice. Classroom quizzing is an effective way to implement retrieval practice, but there are likely a variety of ways to encourage retrieval in educational contexts that have yet to be explored.

Acknowledgments

The writing of this chapter, and some of the research described, was supported in part by grants from the National Science Foundation (DRL-1149363 and DUE-1245476), the Institute of Education Sciences in the U.S. Department of Education (R305A110903 and R305A150546), and the James S. McDonnell Foundation (220020483). The opinions expressed are those of the author and do not represent the views of the National Science Foundation, the Institute of Education Sciences, the U.S. Department of Education, or the James S. McDonnell Foundation.

See also: 2.02 Encoding–Retrieval Interactions. 2.20 Memory for Text and Discourse: Retrieval and Comprehension. 2.26 Spacing Effects on Learning and Memory.

References

Abbott, E.E., 1909. On the analysis of the factors of recall in the learning process. Psychol. Monogr. 11, 159–177. http://dx.doi.org/10.1037/h0093018.

Agarwal, P.K., D'Antonio, L., Roediger III, H.L., McDermott, K.B., McDaniel, M.A., 2014. Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. J. Appl. Res. Mem. Cogn. 3 (3), 131–139. http://dx.doi.org/10.1016/j.jarmac.2014.07.002.

Agarwal, P.K., Finley, J.R., Rose, N.S., Roediger, H.L., 2016. Benefits from retrieval practice are greater for students with lower working memory capacity. Memory 17, 1–8 [Epub ahead of print].

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. Appl. Cogn. Psychol. 22 (7), 861–876. http://dx.doi.org/10.1002/acp.1391.

Agarwal, P.K., Roediger, H.L., 2011. Expectancy of an open-book test decreases performance on a delayed closed-book test. Memory 19 (8), 836–852. http://dx.doi.org/10.1080/09658211.2011.613840.

Anderson, L.S., Healy, A.F., Kole, J.A., Bourne, L.E., 2013. The clicker technique: cultivating efficient teaching and successful learning. Appl. Cogn. Psychol. 27 (2), 222–234. http://dx.doi.org/10.1002/acp.2899.

Aslan, A., Bauml, K.T., 2016. Testing enhances subsequent learning in older but not in younger elementary school children. Dev. Sci. 19 (6), 992–998. http://dx.doi.org/10.1111/desc.12340.

Atkinson, R.C., 1975. Mnemotechnics in second-language learning. Am. Psychol. 30 (8), 821–828. http://dx.doi.org/10.1037/h0077029.

Balota, D.A., Duchek, J.M., Logan, J.M., 2007. Is expanded retrieval practice a superior form of spaced Retrieval? A critical review of the extant literature. In: Nairne, J.S. (Ed.), The Foundations of Remembering: Essays in Honor of Henry L. Roediger, III. Psychology Press, New York, NY, pp. 83–105.

Balota, D.A., Duchek, J.M., Sergent-Marshall, S.D., Roediger, H.L., 2006. Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. Psychol. Aging 21 (1), 19–31. http://dx.doi.org/10.1037/0882-7974.21.1.19.

Bangert-Drowns, R.L., Kulik, J.A., Kulik, C.-I. C., 1991. Effects of frequent classroom testing. J. Educ. Res. 85 (2), 89–99. http://dx.doi.org/10.1080/00220671.1991.10702818.

Batsell, W.R., Perry, J.L., Hanley, E., Hostetter, A.B., 2017. Ecological validity of the testing effect: The use of daily quizzes in introductory psychology. Teach. Psychol. 44 (1), 18–23. http://dx.doi.org/10.1177/0098628316677492.

Beilock, S.L., Kulp, C.A., Holt, L.E., Carr, T.H., 2004. More on the fragility of performance: choking under pressure in mathematical problem solving. J. Exp. Psychol. Gen. 133 (4), 584–600. http://dx.doi.org/10.1037/0096-3445.133.4.584.

Bishara, A.J., Jacoby, L.L., 2008. Aging, spaced retrieval, and inflexible memory performance. Psychon. Bull. Rev. 15 (1), 52–57. http://dx.doi.org/10.3758/pbr.15.1.52.

Bjork, E.L., Bjork, R.A., 2011. Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning. In: Gernsbacher, M.A., Pew, R.W., Hough, L.M., Pomerantz, J.R. (Eds.), Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society. Worth Publishers, New York, NY, pp. 56–64.

Bjork, R.A., 1975. Retrieval as a memory modifier: an interpretation of negative recency and related phenomena. In: Solso, R.L. (Ed.), Information Processing and Cognition: The Loyola Symposium. Erlbaum, Hillsdale, NJ, pp. 123–144.

Bjork, R.A., 1994. Memory and metamemory considerations in the training of human beings. In: Metcalfe, J., Shimamura, A.P. (Eds.), Metacognition: Knowing about Knowing. The MIT Press, Cambridge, MA, pp. 185–205.

Bjork, R.A., 1999. Assessing our own competence: heuristics and illusions. In: Gopher, D., Koriat, A. (Eds.), Attention and Performance. The MIT Press, Cambridge, MA, pp. 435–459.

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

Black, P., Wiliam, D., 2009. Developing the theory of formative assessment. Educ. Assess. Eval. Account. 21 (1), 5–31. http://dx.doi.org/10.1007/s11092-008-9068-5.

Blunt, J.R., Karpicke, J.D., 2014. Learning with retrieval-based concept mapping. J. Educ. Psychol. 106 (3), 849–858. http://dx.doi.org/10.1037/a0035934.

Bouwmeester, S., Verkoeijen, P.P.J.L., 2011. Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. J. Mem. Lang. 65 (1), 32–41. http://dx.doi.org/10.1016/j.jml.2011.02.005.

Brewer, G.A., Marsh, R.L., Meeks, J.T., Clark-Foos, A., Hicks, J.L., 2010. The effects of free recall testing on subsequent source memory. Memory 18 (4), 385–393. http://dx.doi.org/10.1080/09658211003702163.

Brewer, G.A., Unsworth, N., 2012. Individual differences in the effects of retrieval from long-term memory. J. Mem. Lang. 66 (3), 407–415. http://dx.doi.org/10.1016/j.jml.2011.12.009.

Butler, A.C., 2010. Repeated testing produces superior transfer of learning relative to repeated studying. J. Exp. Psychol. Learn. Mem. Cogn. 36 (5), 1118–1133. http://dx.doi.org/10.1037/a0019902.

Butler, A.C., Karpicke, J.D., Roediger, H.L., 2007. The effect of type and timing of feedback on learning from multiple-choice tests. J. Exp. Psychol. Appl. 13 (4), 273–281. http://dx.doi.org/10.1037/1076-898x.13.4.273.

Butler, A.C., Marsh, E.J., Goode, M.K., Roediger III, H.L., 2006. When additional multiple-choice lures aid versus hinder later memory. Appl. Cogn. Psychol. 20 (7), 941–956. http://dx.doi.org/10.1002/acp.1239.

Butler, A.C., Marsh, E.J., Slavinsky, J.P., Baraniuk, R.G., 2014. Integrating cognitive science and technology improves learning in a STEM classroom. Educ. Psychol. Rev. 26 (2), 331–340. http://dx.doi.org/10.1007/s10648-014-9256-4.

Butler, A.C., Roediger, H.L., 2007. Testing improves long-term retention in a simulated classroom setting. Eur. J. Cogn. Psychol. 19 (4–5), 514–527. http://dx.doi.org/10.1080/09541440701326097.

Butler, A.C., Roediger, H.L., 2008. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Mem. Cogn. 36 (3), 604–616. http://dx.doi.org/10.3758/mc.36.3.604.

Caldwell, J.E., 2007. Clickers in the large classroom: current research and best-practice tips. CBE Life Sci. Educ. 6 (1), 9–20. http://dx.doi.org/10.1187/cbe.06-12-0205.

Campbell, J., Mayer, R.E., 2009. Questioning as an instructional method: does it affect learning from lectures? Appl. Cogn. Psychol. 23 (6), 747–759. http://dx.doi.org/10.1002/acp.1513.

Carpenter, S.K., 2009. Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. J. Exp. Psychol. Learn. Mem. Cogn. 35 (6), 1563–1569. http://dx.doi.org/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. J. Exp. Psychol. Learn. Mem. Cogn. 37 (6), 1547–1552. http://dx.doi.org/10.1037/a0024140.

Carpenter, S.K., 2012. Testing enhances the transfer of learning. Curr. Dir. Psychol. Sci. 21 (5), 279–283. http://dx.doi.org/10.1177/0963721412452728.

Carpenter, S.K., DeLosh, E.L., 2005. Application of the testing and spacing effects to name learning. Appl. Cogn. Psychol. 19 (5), 619–636. http://dx.doi.org/10.1002/acp.1101.

Carpenter, S.K., DeLosh, E.L., 2006. Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. Mem. Cogn. 34 (2), 268–276. http://dx.doi.org/10.3758/bf03193405.

Carpenter, S.K., Kelly, J.W., 2012. Tests enhance retention and transfer of spatial learning. Psychon. Bull. Rev. 19 (3), 443–448. http://dx.doi.org/10.3758/s13423-012-0221-2.

Carpenter, S.K., Lund, T.J.S., Coffman, C.R., Armstrong, P.I., Lamm, M.H., Reason, R.D., 2016. A classroom study on the relationship between student achievement and retrieval-enhanced learning. Educ. Psychol. Rev. 28 (2), 353–375. http://dx.doi.org/10.1007/s10648-015-9311-9.

Carpenter, S.K., Pashler, H., 2007. Testing beyond words: using tests to enhance visuospatial map learning. Psychon. Bull. Rev. 14 (3), 474–478. http://dx.doi.org/10.3758/bf03194092.

Carpenter, S.K., Pashler, H., Cepeda, N.J., 2009. Using tests to enhance 8th grade students' retention of U.S. history facts. Appl. Cogn. Psychol. 23 (6), 760–771. http://dx.doi.org/10.1002/acp.1507.

Carpenter, S.K., Pashler, H., Vul, E., 2006. What types of learning are enhanced by a cued recall test? Psychon. Bull. Rev. 13 (5), 826–830. http://dx.doi.org/10.3758/bf03194004.

Carpenter, S.K., Pashler, H., Wixted, J.T., Vul, E., 2008. The effects of tests on learning and forgetting. Mem. Cogn. 36 (2), 438–448. http://dx.doi.org/10.3758/mc.36.2.438.

Carpenter, S.K., Yeung, K.L., 2017. The role of mediator strength in learning from retrieval. J. Mem. Lang. 92, 128–141. http://dx.doi.org/10.1016/j.jml.2016.06.008.

Carrier, M., Pashler, H., 1992. The influence of retrieval on retention. Mem. Cogn. 20 (6), 633–642. http://dx.doi.org/10.3758/bf03202713.

Chan, J.C.K., 2009. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. J. Mem. Lang. 61 (2), 153–170. http://dx.doi.org/10.1016/j.jml.2009.04.004.

Chan, J.C.K., 2010. Long-term effects of testing on the recall of nontested materials. Memory 18 (1), 49–57. http://dx.doi.org/10.1080/09658210903405737.

Chan, J.C.K., McDermott, K.B., 2007. The testing effect in recognition memory: a dual process account. J. Exp. Psychol. Learn. Mem. Cogn. 33 (2), 431–437. http://dx.doi.org/10.1037/0278-7393.33.2.431.

Chan, J.C.K., McDermott, K.B., Roediger III, H.L., 2006. Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. J. Exp. Psychol. Gen. 135 (4), 553–571. http://dx.doi.org/10.1037/0096-3445.135.4.553.

Coane, J.H., 2013. Retrieval practice and elaborative encoding benefit memory in younger and older adults. J. Appl. Res. Mem. Cogn. 2 (2), 95–100. http://dx.doi.org/10.1016/j.jarmac.2013.04.001.

Congleton, A., Rajaram, S., 2012. The origin of the interaction between learning method and delay in the testing effect: the roles of processing and conceptual retrieval organization. Mem. Cogn. 40 (4), 528–539. http://dx.doi.org/10.3758/s13421-011-0168-y.

Coppens, L.C., Verkoeijen, P.P.J.L., Bouwmeester, S., Rikers, R.M.J.P., 2016. The testing effect for mediator final test cues and related final test cues in online and laboratory experiments. BMC Psychol. 4 (25), 1–14. http://dx.doi.org/10.1186/s40359-016-0127-2.

Coppens, L.C., Verkoeijen, P.P.J.L., Rikers, R.M.J.P., 2011. Learning Adinkra symbols: the effect of testing. J. Cogn. Psychol. 23 (3), 351–357. http://dx.doi.org/10.1080/20445911.2011.507188.

Cull, W.L., 2000. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. Appl. Cogn. Psychol. 14 (3), 215–235. http://dx.doi.org/10.1002/(sici)1099-0720(200005/06)14:3<215::aid-acp640>3.0.co;2–1.

de Jonge, M., Tabbers, H.K., Rikers, R.M.J.P., 2015. The effect of testing on the retention of coherent and incoherent text material. Educ. Psychol. Rev. 27 (2), 305–315. http://dx.doi.org/10.1007/s10648-015-9300-z.

Eglington, L.G., Kang, S.H.K., 2016. Retrieval practice benefits deductive inference. Educ. Psychol. Rev. http://dx.doi.org/10.1007/s10648-016-9386-y.

Finley, J.R., Benjamin, A.S., Hays, M.J., Bjork, R.A., Kornell, N., 2011. Benefits of accumulating versus diminishing cues in recall. J. Mem. Lang. 64 (4), 289–298. http://dx.doi.org/10.1016/j.jml.2011.01.006.

Frase, L.T., 1968. Effect of question location, pacing, and mode upon retention of prose material. J. Educ. Psychol. 59 (4), 244–249. http://dx.doi.org/10.1037/h0025947.

Friedman, R.B., Sullivan, K.L., Snider, S.F., Luta, G., Jones, K.T., 2017. Leveraging the test effect to improve maintenance of the gains achieved through cognitive rehabilitation. Neuropsychology 31 (2), 220–228. http://dx.doi.org/10.1037/neu0000318.

Fritz, C.O., Morris, P.E., Nolan, D., Singleton, J., 2007. Expanding retrieval practice: an effective aid to preschool children's learning. Q. J. Exp. Psychol. 60 (7), 991–1004. http://dx.doi.org/10.1080/17470210600823595.

Gates, A.I., 1917. Recitation as a factor in memorizing. Arch. Psychol. 6 (40), 1–104.

Glover, J.A., 1989. The 'testing' phenomenon: not gone but nearly forgotten. J. Educ. Psychol. 81 (3), 392–399. http://dx.doi.org/10.1037/0022-0663.81.3.392.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H.K., 2014a. The effect of retrieval practice in primary school vocabulary learning. Appl. Cogn. Psychol. 28 (1), 135–142. http://dx.doi.org/10.1002/acp.2956.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H.K., Zwaan, R.A., 2014b. The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. J. Appl. Res. Mem. Cogn. 3 (3), 177–182. http://dx.doi.org/10.1016/j.jarmac.2014.05.003.

Grimaldi, P.J., Karpicke, J.D., 2014. Guided retrieval practice of educational materials using automated scoring. J. Educ. Psychol. 106 (1), 58–68. http://dx.doi.org/10.1037/a0033208.

Halamish, V., Bjork, R.A., 2011. When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. J. Exp. Psychol. Learn. Mem. Cogn. 37 (4), 801–812. http://dx.doi.org/10.1037/a0023219.

Hanawalt, N.G., 1937. Memory trace for figures in recall and recognition. Arch. Psychol. 31 (216), 5–89.

Henkel, L.A., 2007. The benefits and costs of repeated memory tests for young and older adults. Psychol. Aging 22 (3), 580–595. http://dx.doi.org/10.1037/0882-7974.22.3.580.

Henkel, L.A., 2008. Maximizing the benefits and minimizing the costs of repeated memory tests for older adults. Psychol. Aging 23 (2), 250–262. http://dx.doi.org/10.1037/0882-7974.23.2.250.

Hills, T.T., Todd, P.M., Jones, M.N., 2015. Foraging in semantic fields: how we search through memory. Top. Cogn. Sci. 7 (3), 513–534. http://dx.doi.org/10.1111/tops.12151.

Hinze, S.R., Rapp, D.N., 2014. Retrieval (sometimes) enhances learning: performance pressure reduces the benefits of retrieval practice. Appl. Cogn. Psychol. 28 (4), 597–606. http://dx.doi.org/10.1002/acp.3032.

Hinze, S.R., Wiley, J., 2011. Testing the limits of testing effects using completion tests. Memory 19 (3), 290–304. http://dx.doi.org/10.1080/09658211.2011.560121.

Hinze, S.R., Wiley, J., Pellegrino, J.W., 2013. The importance of constructive comprehension processes in learning from tests. J. Mem. Lang. 69 (2), 151–164. http://dx.doi.org/10.1016/j.jml.2013.03.002.

Hogan, R.M., Kintsch, W., 1971. Differential effects of study and test trials on long-term recognition and recall. J. Verbal Learn. Verbal Behav. 10 (5), 562–567. http://dx.doi.org/10.1016/s0022-5371(71)80029-4.

Hopkins, R.F., Lyle, K.B., Hieb, J.L., Ralston, P.A.S., 2016. Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. Educational Psychology Review 28 (4), 853–873. http://dx.doi.org/10.1007/s10648-015-9349-8.

Howard, M.W., Kahana, M.J., 2002. A distributed representation of temporal context. J. Math. Psychol. 46 (3), 269–299. http://dx.doi.org/10.1006/jmps.2001.1388.

Jacoby, L.L., 1991. A process dissociation framework: separating automatic from intentional uses of memory. J. Mem. Lang. 30 (5), 513–541. http://dx.doi.org/10.1016/0749-596x(91)90025-f.

Jaeger, A., Eisenkraemer, R.E., Stein, L.M., 2015. Test-enhanced learning in third-grade children. Educ. Psychol. 35 (4), 513–521. http://dx.doi.org/10.1080/01443410.2014.963030.

Jensen, J.L., McDaniel, M.A., Woodard, S.M., Kummer, T.A., 2014. Teaching to the test … or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. Educational Psychology Review 26 (2), 307–329. http://dx.doi.org/10.1007/s10648-013-9248-9.

Johnson, C.I., Mayer, R.E., 2009. A testing effect with multimedia learning. J. Educ. Psychol. 101 (3), 621–629. http://dx.doi.org/10.1037/a0015183.

Kang, S.H.K., 2010. Enhancing visuospatial learning: the benefit of retrieval practice. Mem. Cogn. 38 (8), 1009–1017. http://dx.doi.org/10.3758/mc.38.8.1009.

Kang, S.H.K., Lindsey, R.V., Mozer, M.C., Pashler, H., 2014. Retrieval practice over the long term: should spacing be expanding or equal-interval? Psychon. Bull. Rev. 21 (6), 1544–1550. http://dx.doi.org/10.3758/s13423-014-0636-z.

Kang, S.H.K., McDermott, K.B., Roediger, H.L., 2007. Test format and corrective feedback modify the effect of testing on long-term retention. Eur. J. Cogn. Psychol. 19 (4–5), 528–558. http://dx.doi.org/10.1080/09541440601056620.

Karpicke, J.D., 2009. Metacognitive control and strategy selection: deciding to practice retrieval during learning. J. Exp. Psychol. Gen. 138 (4), 469–486. http://dx.doi.org/10.1037/a0017341.

Karpicke, J.D., 2012. Retrieval-based learning: active retrieval promotes meaningful learning. Curr. Dir. Psychol. Sci. 21 (3), 157–163. http://dx.doi.org/10.1177/0963721412443552.

Karpicke, J.D., 2017. Concept mapping. In: B. Frey (Ed.), The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. Sage Publications, (in press).

Karpicke, J.D., Aue, W.R., 2015. The testing effect Is alive and well with complex materials. Educ. Psychol. Rev. 27 (2), 317–326. http://dx.doi.org/10.1007/s10648-015-9309-3.

Karpicke, J.D., Bauernschmidt, A., 2011. Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. J. Exp. Psychol. Learn. Mem. Cogn. 37 (5), 1250–1257. http://dx.doi.org/10.1037/a0023436.

Karpicke, J.D., Blunt, J.R., 2011. Retrieval practice produces more learning than elaborative studying with concept mapping. Science 331 (6018), 772–775. http://dx.doi.org/10.1126/science.1199327.

Karpicke, J.D., Blunt, J.R., Smith, M.A., 2016. Retrieval-based learning: positive effects of retrieval practice in elementary school children. Front. Psychol. 7 (350), 1–8. http://dx.doi.org/10.3389/fpsyg.2016.00350.

Karpicke, J.D., Blunt, J.R., Smith, M.A., Karpicke, S.S., 2014a. Retrieval-based learning: the need for guided retrieval in elementary school children. J. Appl. Res. Mem. Cogn. 3 (3), 198–206. http://dx.doi.org/10.1016/j.jarmac.2014.07.008.

Karpicke, J.D., Lehman, M., Aue, W.R., 2014b. Retrieval-based learning: an episodic context account. In: Ross, B.H. (Ed.), Psychology of Learning and Motivation, vol. 61. Elsevier Academic Press, San Diego, CA, pp. 237–284.

Karpicke, J.D., Roediger, H.L., 2007a. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. J. Exp. Psychol. Learn. Mem. Cogn. 33 (4), 704–719. http://dx.doi.org/10.1037/0278-7393.33.4.704.

Karpicke, J.D., Roediger, H.L., 2007b. Repeated retrieval during learning is the key to long-term retention. J. Mem. Lang. 57 (2), 151–162. http://dx.doi.org/10.1016/j.jml.2006.09.004.

Karpicke, J.D., Roediger, H.L., 2008. The critical importance of retrieval for learning. Science 319 (5865), 966–968. http://dx.doi.org/10.1126/science.1152408.

Karpicke, J.D., Roediger, H.L., 2010. Is expanding retrieval a superior method for learning text materials? Mem. Cogn. 38 (1), 116–124. http://dx.doi.org/10.3758/mc.38.1.116.

Karpicke, J.D., Smith, M.A., 2012. Separate mnemonic effects of retrieval practice and elaborative encoding. J. Mem. Lang. 67 (1), 17–29. http://dx.doi.org/10.1016/j.jml.2012.02.004.

Karpicke, J.D., Zaromb, F.M., 2010. Retrieval mode distinguishes the testing effect from the generation effect. J. Mem. Lang. 62 (3), 227–239. http://dx.doi.org/10.1016/j.jml.2009.11.010.

Kelly, J.W., Carpenter, S.K., Sjolund, L.A., 2015. Retrieval enhances route knowledge acquisition, but only when movement errors are prevented. J. Exp. Psychol. Learn. Mem. Cogn. 41 (5), 1540–1547. http://dx.doi.org/10.1037/a0038685 10.1037/a0038685.supp.

Keresztes, A., Kaiser, D., Kovacs, G., Racsmany, M., 2014. Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. Cereb. Cortex 24 (11), 3025–3035. http://dx.doi.org/10.1093/cercor/bht158.

Knouse, L.E., Rawson, K.A., Vaughn, K.E., Dunlosky, J., 2016. Does testing improve learning for college students with attention-deficit/hyperactivity disorder? Clin. Psychol. Sci. 4 (1), 136–143. http://dx.doi.org/10.1177/2167702614565175.

Kolers, P.A., Roediger, H.L., 1984. Procedures of mind. J. Verbal Learn. Verbal Behav. 23 (4), 425–449. http://dx.doi.org/10.1016/s0022-5371(84)90282-2.

Kornell, N., Bjork, R.A., Garcia, M.A., 2011. Why tests appear to prevent forgetting: a distribution-based bifurcation model. J. Mem. Lang. 65 (2), 85–97. http://dx.doi.org/10.1016/j.jml.2011.04.002.

Kornell, N., Klein, P.J., Rawson, K.A., 2015. Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. J. Exp. Psychol. Learn. Mem. Cogn. 41 (1), 283–294. http://dx.doi.org/10.1037/a0037850.

Kupper-Tetzel, C.E., Kapler, I.V., Wiseheart, M., 2014. Contracting, equal, and expanding learning schedules: the optimal distribution of learning sessions depends on retention interval. Mem. Cogn. 42 (5), 729–741. http://dx.doi.org/10.3758/s13421-014-0394-1.

Landauer, T.K., Bjork, R.A., 1978. Optimum rehearsal patterns and name learning. In: Gruneberg, M.M., Morris, P.E., Sykes, R.N. (Eds.), Practical Aspects of Memory. Academic Press, London, pp. 625–632.

Larsen, D.P., Butler, A.C., Lawson, A.L., Roediger III, H.L., 2013. The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. Adv. Health Sci. Educ. 18 (3), 409–425. http://dx.doi.org/10.1007/s10459-012-9379-7.

Leahy, W., Hanham, J., Sweller, J., 2015. High element interactivity information during problem solving may lead to failure to obtain the testing effect. Educ. Psychol. Rev. 27 (2), 291–304. http://dx.doi.org/10.1007/s10648-015-9296-4.

Lechuga, M.T., Ortega-Tudela, J.M., Gomez-Ariza, C.J., 2015. Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. Learn. Instr. 40, 61–68. http://dx.doi.org/10.1016/j.learninstruc.2015.08.002.

Leeming, F.C., 2002. The exam-a-day procedure improves performance in psychology classes. Teach. Psychol. 29 (3), 210–212. http://dx.doi.org/10.1207/s15328023top2903_06.

Lehman, M., Karpicke, J.D., 2016. Elaborative retrieval: do semantic mediators improve memory? J. Exp. Psychol. Learn. Mem. Cogn. 42 (10), 1573–1591. http://dx.doi.org/10.1037/xlm0000267.

Lehman, M., Malmberg, K.J., 2013. A buffer model of memory encoding and temporal correlations in retrieval. Psychol. Rev. 120 (1), 155–189. http://dx.doi.org/10.1037/a0030851.

Lehman, M., Smith, M.A., Karpicke, J.D., 2014. Toward an episodic context account of retrieval-based learning: dissociating retrieval practice and elaboration. J. Exp. Psychol. Learn. Mem. Cogn. 40 (6), 1787–1794. http://dx.doi.org/10.1037/xlm0000012.

Lindsey, R.V., Shroyer, J.D., Pashler, H., Mozer, M.C., 2014. Improving students' long-term knowledge retention through personalized review. Psychol. Sci. 25 (3), 639–647. http://dx.doi.org/10.1177/0956797613504302.

Lipko-Speed, A., Dunlosky, J., Rawson, K.A., 2014. Does testing with feedback help grade-school children learn key concepts in science? J. Appl. Res. Mem. Cogn. 3 (3), 171–176. http://dx.doi.org/10.1016/j.jarmac.2014.04.002.

Lipowski, S.L., Pyc, M.A., Dunlosky, J., Rawson, K.A., 2014. Establishing and explaining the testing effect in free recall for young children. Dev. Psychol. 50 (4), 994–1000. http://dx.doi.org/10.1037/a0035202.

Little, J.L., Bjork, E.L., 2015. Optimizing multiple-choice tests as tools for learning. Mem. Cogn. 43 (1), 14–26. http://dx.doi.org/10.3758/s13421-014-0452-8.

Little, J.L., Bjork, E.L., Bjork, R.A., Angello, G., 2012. Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. Psychol. Sci. 23 (11), 1337–1344. http://dx.doi.org/10.1177/0956797612443370.

Little, J.L., Storm, B.C., Bjork, E.L., 2011. The costs and benefits of testing text materials. Memory 19 (4), 346–359. http://dx.doi.org/10.1080/09658211.2011.569725.

Logan, J.M., Balota, D.A., 2008. Expanded vs. equal interval spaced retrieval practice: exploring different schedules of spacing and retention interval in younger and older adults. Aging Neuropsychol. Cogn. 15 (3), 257–280. http://dx.doi.org/10.1080/13825580701322171.

Lyle, K.B., Crawford, N.A., 2011. Retrieving essential material at the end of lectures improves performance on statistics exams. Teach. Psychol. 38 (2), 94–97. http://dx.doi.org/10.1177/0098628311401587.

Maddox, G.B., Balota, D.A., Coane, J.H., Duchek, J.M., 2011. The role of forgetting rate in producing a benefit of expanded over equal spaced retrieval in young and older adults. Psychol. Aging 26 (3), 661–670. http://dx.doi.org/10.1037/a0022942.

Mandler, G., Rabinowitz, J.C., 1981. Appearance and reality: does a recognition test really improve subsequent recall and recognition? J. Exp. Psychol. Hum. Learn. Mem. 7 (2), 79–90. http://dx.doi.org/10.1037/0278-7393.7.2.79.

Marsh, E.J., Agarwal, P.K., Roediger III, H.L., 2009. Memorial consequences of answering SAT II questions. J. Exp. Psychol. Appl. 15 (1), 1–11. http://dx.doi.org/10.1037/a0014721.

Marsh, E.J., Cantor, A.D., 2014. Learning from the test: do's and don'ts for using multiple-choice tests. In: McDaniel, M.A., Frey, R.F., Fitzpatrick, S.M., Roediger, H.L. (Eds.), Integrating Cognitive Science with Innovative Teaching in STEM Disciplines. http://dx.doi.org/10.7936/K7Z60KZK.

Marsh, E.J., Fazio, L.K., Goswick, A.E., 2012. Memorial consequences of testing school-aged children. Memory 20 (8), 899–906. http://dx.doi.org/10.1080/09658211.2012.708757.

Martyn, M., 2007. Clickers in the classroom: an active learning approach. Educ. Q. 2, 71–74.

Mayer, R.E., 2008. Learning and Instruction, second ed. Pearson, Upper Saddle River, NJ.

Mayer, R.E., Gallini, J.K., 1990. When is an illustration worth ten thousand words? J. Educ. Psychol. 82 (4), 715–726. http://dx.doi.org/10.1037/0022-0663.82.4.715.

Mayer, R.E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Zhang, H., 2009. Clickers in college classrooms: fostering learning with questioning methods in large lecture classes. Contemp. Educ. Psychol. 34 (1), 51–57. http://dx.doi.org/10.1016/j.cedpsych.2008.04.002.

McDaniel, M.A., 2007. Rediscovering transfer as a central concept. In: Roediger, H.L., Dudai, Y., Fitzpatrick, S.M. (Eds.), Science of Memory: Concepts. Oxford University Press, New York, pp. 355–360.

McDaniel, M.A., Agarwal, P.K., Huelser, B.J., McDermott, K.B., Roediger III, H.L., 2011. Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. J. Educ. Psychol. 103 (2), 399–414. http://dx.doi.org/10.1037/a0021782.

McDaniel, M.A., Anderson, J.L., Derbish, M.H., Morrisette, N., 2007a. Testing the testing effect in the classroom. Eur. J. Cogn. Psychol. 19 (4–5), 494–513. http://dx.doi.org/10.1080/09541440701326154.
McDaniel, M.A., Bugg, J.M., 2008. Instability in memory phenomena: a common puzzle and a unifying explanation. Psychon. Bull. Rev. 15 (2), 237–255. http://dx.doi.org/10.3758/pbr.15.2.237.
McDaniel, M.A., Butler, A.C., 2011. A contextual framework for understanding when difficulties are desirable. In: Benjamin, A.S. (Ed.), Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork. Psychology Press, New York, NY, pp. 175–198.
McDaniel, M.A., Einstein, G.O., 1989. Material-appropriate processing: a contextualist approach to reading and studying strategies. Educ. Psychol. Rev. 1 (2), 113–145. http://dx.doi.org/10.1007/BF01326639.
McDaniel, M.A., Howard, D.C., Einstein, G.O., 2009. The read-recite-review study strategy: effective and portable. Psychol. Sci. 20 (4), 516–522. http://dx.doi.org/10.1111/j.1467-9280.2009.02325.x.
McDaniel, M.A., Roediger III, H.L., McDermott, K.B., 2007b. Generalizing test-enhanced learning from the laboratory to the classroom. Psychon. Bull. Rev. 14 (2), 200–206. http://dx.doi.org/10.3758/bf03194052.
McDaniel, M.A., Thomas, R.C., Agarwal, P.K., McDermott, K.B., Roediger, H.L., 2013. Quizzing in middle-school science: successful transfer performance on classroom exams. Appl. Cogn. Psychol. 27 (3), 360–372. http://dx.doi.org/10.1002/acp.2914.
McDaniel, M.A., Wildman, K.M., Anderson, J.L., 2012. Using quizzes to enhance summative-assessment performance in a web-based class: an experimental study. J. Appl. Res. Mem. Cogn. 1 (1), 18–26. http://dx.doi.org/10.1016/j.jarmac.2011.10.001.
McDermott, K.B., Agarwal, P.K., D'Antonio, L., Roediger III, H.L., McDaniel, M.A., 2014. Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. J. Exp. Psychol. Appl. 20 (1), 3–21. http://dx.doi.org/10.1037/xap0000004.
Meyer, A.N.D., Logan, J.M., 2013. Taking the testing effect beyond the college freshman: benefits for lifelong learning. Psychol. Aging 28 (1), 142–147. http://dx.doi.org/10.1037/a003089010.1037/a0030890.supp.
Middleton, E.L., Schwartz, M.F., Rawson, K.A., Garvey, K., 2015. Test-enhanced learning versus errorless learning in aphasia rehabilitation: testing competing psychological principles. J. Exp. Psychol. Learn. Mem. Cogn. 41 (4), 1253–1261. http://dx.doi.org/10.1037/xlm0000091.
Middleton, E.L., Schwartz, M.F., Rawson, K.A., Traut, H., Verkuilen, J., 2016. Towards a theory of learning for naming rehabilitation: retrieval practice and spacing effects. J. Speech Lang. Hear Res. 59 (5), 1111–1122. http://dx.doi.org/10.1044/2016_JSLHR-L-15-0303.
Mok, W.S.Y., Chan, W.W.L., 2016. How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? Instr. Sci. 44 (6), 567–581. http://dx.doi.org/10.1007/s11251-016-9393-x.
Morris, C.D., Bransford, J.D., Franks, J.J., 1977. Levels of processing versus transfer appropriate processing. J. Verbal Learn. Verbal Behav. 16 (5), 519–533. http://dx.doi.org/10.1016/s0022-5371(77)80016-9.
Mulligan, N.W., Peterson, D.J., 2015. Negative and positive testing effects in terms of item-specific and relational information. J. Exp. Psychol. Learn. Mem. Cogn. 41 (3), 859–871. http://dx.doi.org/10.1037/xlm0000056.
Nairne, J.S., Riegler, G.L., Serra, M., 1991. Dissociative effects of generation on item and order retention. J. Exp. Psychol. Learn. Mem. Cogn. 17 (4), 702–709. http://dx.doi.org/10.1037/0278-7393.17.4.702.
Pan, S.C., Pashler, H., Potter, Z.E., Rickard, T.C., 2015. Testing enhances learning across a range of episodic memory abilities. J. Mem. Lang. 83, 53–61. http://dx.doi.org/10.1016/j.jml.2015.04.001.
Park, J., 2005. Learning in a new computerized testing system. J. Educ. Psychol. 97 (3), 436–443. http://dx.doi.org/10.1037/0022-0663.97.3.436.
Pastotter, B., Weber, J., Bauml, K.-H.T., 2013. Using testing to improve learning after severe traumatic brain injury. Neuropsychology 27 (2), 280–285. http://dx.doi.org/10.1037/a0031797.
Pennebaker, J.W., Gosling, S.D., Ferrell, J.D., 2013. Daily online testing in large classes: boosting college performance while reducing achievement gaps. PLoS One 8 (11), e79774. http://dx.doi.org/10.1371/journal.pone.0079774.
Peterson, D.J., Mulligan, N.W., 2013. The negative testing effect and multifactor account. J. Exp. Psychol. Learn. Mem. Cogn. 39 (4), 1287–1293. http://dx.doi.org/10.1037/a0031337.
Pu, X., Tse, C.-S., 2014. The influence of intentional versus incidental retrieval practices on the role of recollection in test-enhanced learning. Cogn. Process. 15 (1), 55–64. http://dx.doi.org/10.1007/s10339-013-0580-2.
Pyc, M.A., Rawson, K.A., 2007. Examining the efficiency of schedules of distributed retrieval practice. Mem. Cogn. 35 (8), 1917–1927. http://dx.doi.org/10.3758/bf03192925.
Pyc, M.A., Rawson, K.A., 2009. Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? J. Mem. Lang. 60 (4), 437–447. http://dx.doi.org/10.1016/j.jml.2009.01.004.
Pyc, M.A., Rawson, K.A., 2010. Why testing improves memory: mediator effectiveness hypothesis. Science 330 (6002), 335. http://dx.doi.org/10.1126/science.1191465.
Raaijmakers, J.G.W., Shiffrin, R.M., 1981. Search of associative memory. Psychol. Rev. 88 (2), 93–134. http://dx.doi.org/10.1037/0033-295x.88.2.93.
Rajaram, S., 1993. Remembering and knowing: two means of access to the personal past. Mem. Cogn. 21 (1), 89–102. http://dx.doi.org/10.3758/bf03211168.
Rawson, K.A., 2015. The status of the testing effect for complex materials: still a winner. Educ. Psychol. Rev. 27 (2), 327–331. http://dx.doi.org/10.1007/s10648-015-9308-4.
Rawson, K.A., Vaughn, K.E., Carpenter, S.K., 2015a. Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. Mem. Cogn. 43 (4), 619–633. http://dx.doi.org/10.3758/s13421-014-0477-z.
Rawson, K.A., Wissman, K.T., Vaughn, K.E., 2015b. Does testing impair relational processing? Failed attempts to replicate the negative testing effect. J. Exp. Psychol. Learn. Mem. Cogn. 41 (5), 1326–1336. http://dx.doi.org/10.1037/xlm0000127.
Roediger, H.L., Agarwal, P.K., McDaniel, M.A., McDermott, K.B., 2011. Test-enhanced learning in the classroom: long-term improvements from quizzing. J. Exp. Psychol. Appl. 17 (4), 382–395. http://dx.doi.org/10.1037/a0026252.
Roediger, H.L., Karpicke, J.D., 2006a. The power of testing memory: basic research and implications for educational practice. Perspect. Psychol. Sci. 1 (3), 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x.
Roediger, H.L., Karpicke, J.D., 2006b. Test-enhanced learning: taking memory tests improves long-term retention. Psychol. Sci. 17 (3), 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x.
Roediger, H.L., Karpicke, J.D., 2011. Intricacies of spaced retrieval: a resolution. In: Benjamin, A.S. (Ed.), Successful Remembering and Successful Forgetting: Essays in Honor of Robert A. Bjork. Psychology Press, New York, pp. 23–48.
Roediger, H.L., Karpicke, J.D., 2017. Reflections on the resurgence of interest in the testing effect. Perspect. Psychol. Sci. (in press).
Roediger, H.L., Marsh, E.J., 2005. The positive and negative consequences of multiple-choice testing. J. Exp. Psychol. Learn. Mem. Cogn. 31 (5), 1155–1159. http://dx.doi.org/10.1037/0278-7393.31.5.1155.
Roediger, H.L., Neely, J.H., 1982. Retrieval blocks in episodic and semantic memory. Can. J. Psychol./Revue Can. de Psychol. 36 (2), 213–242. http://dx.doi.org/10.1037/h0080640.
Roenker, D.L., Thompson, C.P., Brown, S.C., 1971. Comparison of measures for the estimation of clustering in free recall. Psychol. Bull. 76 (1), 45–48. http://dx.doi.org/10.1037/h0031355.
Rogalski, Y., Altmann, L.J.P., Rosenbek, J.C., 2014. Retrieval practice and testing improve memory in older adults. Aphasiology 28 (4), 381–400. http://dx.doi.org/10.1080/02687038.2013.870965.
Rohrer, D., Taylor, K., Sholar, B., 2010. Tests enhance the transfer of learning. J. Exp. Psychol. Learn. Mem. Cogn. 36 (1), 233–239. http://dx.doi.org/10.1037/a0017678.

Rosburg, T., Johansson, M., Weigl, M., Mecklinger, A., 2015. How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. Cogn. Affect. Behav. Neurosci. 15 (1), 195–210. http://dx.doi.org/10.3758/s13415-014-0310-y.

Rowland, C.A., 2014. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. Psychol. Bull. 140 (6), 1432–1463. http://dx.doi.org/10.1037/a0037559.

Rowland, C.A., DeLosh, E.L., 2014. Benefits of testing for nontested information: retrieval-induced facilitation of episodically bound material. Psychon. Bull. Rev. 21 (6), 1516–1523. http://dx.doi.org/10.3758/s13423-014-0625-2.

Rowland, C.A., DeLosh, E.L., 2015. Mnemonic benefits of retrieval practice at short retention intervals. Memory 23 (3), 403–419. http://dx.doi.org/10.1080/09658211.2014.889710.

Sederberg, P.B., Miller, J.F., Howard, M.W., Kahana, M.J., 2010. The temporal contiguity effect predicts episodic memory performance. Mem. Cogn. 38 (6), 689–699. http://dx.doi.org/10.3758/mc.38.6.689.

Shiffrin, R.M., Steyvers, M., 1997. A model for recognition memory: REM–retrieving effectively from memory. Psychon. Bull. Rev. 4 (2), 145–166. http://dx.doi.org/10.3758/bf03209391.

Slamecka, N.J., Katsaiti, L.T., 1988. Normal forgetting of verbal lists as a function of prior testing. J. Exp. Psychol. Learn. Mem. Cogn. 14 (4), 716–727. http://dx.doi.org/10.1037/0278-7393.14.4.716.

Smith, M.A., Blunt, J.R., Whiffen, J.W., Karpicke, J.D., 2016. Does providing prompts during retrieval practice improve learning? Appl. Cogn. Psychol. 30 (4), 544–553. http://dx.doi.org/10.1002/acp.3227.

Smith, M.A., Karpicke, J.D., 2014. Retrieval practice with short-answer, multiple-choice, and hybrid tests. Memory 22 (7), 784–802. http://dx.doi.org/10.1080/09658211.2013.831454.

Smith, M.A., Roediger, H.L., Karpicke, J.D., 2013. Covert retrieval practice benefits retention as much as overt retrieval practice. J. Exp. Psychol. Learn. Mem. Cogn. 39 (6), 1712–1725. http://dx.doi.org/10.1037/a0033569.

Soderstrom, N.C., Bjork, R.A., 2015. Learning versus performance: an integrative review. Perspect. Psychol. Sci. 10 (2), 176–199. http://dx.doi.org/10.1177/1745691615569000.

Storm, B.C., Bjork, R.A., Storm, J.C., 2010. Optimizing retrieval as a learning event: when and why expanding retrieval practice enhances long-term retention. Mem. Cogn. 38 (2), 244–253. http://dx.doi.org/10.3758/mc.38.2.244.

Sumowski, J.F., Chiaravalloti, N., DeLuca, J., 2010a. Retrieval practice improves memory in multiple sclerosis: clinical application of the testing effect. Neuropsychology 24 (2), 267–272. http://dx.doi.org/10.1037/a0017533.

Sumowski, J.F., Wood, H.G., Chiaravalloti, N., Wylie, G.R., Lengenfelder, J., Deluca, J., 2010b. Retrieval practice: a simple strategy for improving memory after traumatic brain injury. J. Int. Neuropsychol. Soc. 16 (6), 1147–1150. http://dx.doi.org/10.1017/s1355617710001128.

Sumowski, J.F., Leavitt, V.M., Cohen, A., Paxton, J., Chiaravalloti, N.D., DeLuca, J., 2013. Retrieval practice is a robust memory aid for memory-impaired patients with MS. Mult. Scler. 19 (14), 1943–1946. http://dx.doi.org/10.1177/1352458513485980.

Thompson, C.P., Wenger, S.K., Bartling, C.A., 1978. How recall facilitates subsequent recall: a reappraisal. J. Exp. Psychol. Hum. Learn. Mem. 4 (3), 210–221. http://dx.doi.org/10.1037/0278-7393.4.3.210.

Tran, R., Rohrer, D., Pashler, H., 2015. Retrieval practice: the lack of transfer to deductive inferences. Psychon. Bull. Rev. 22 (1), 135–140. http://dx.doi.org/10.3758/s13423-014-0646-x.

Tse, C.-S., Balota, D.A., Roediger III, H.L., 2010. The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. Psychol. Aging 25 (4), 833–845. http://dx.doi.org/10.1037/a0019933.

Tse, C.-S., Pu, X., 2012. The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. J. Exp. Psychol. Appl. 18 (3), 253–264. http://dx.doi.org/10.1037/a0029190.

Tulving, E., 1985. Memory and consciousness. Can. Psychol./Psychol. Can. 26 (1), 1–12. http://dx.doi.org/10.1037/h0080017.

Underwood, B.J., 1964. Degree of learning and the measurement of forgetting. J. Verbal Learn. Verbal Behav. 3 (2), 112–129. http://dx.doi.org/10.1016/s0022-5371(64)80028-1.

van den Broek, G.S.E., Segers, E., Takashima, A., Verhoeven, L., 2014. Do testing effects change over time? Insights from immediate and delayed retrieval speed. Memory 22 (7), 803–812. http://dx.doi.org/10.1080/09658211.2013.831455.

van den Broek, G.S.E., Takashima, A., Segers, E., Fernandez, G., Verhoeven, L., 2013. Neural correlates of testing effects in vocabulary learning. NeuroImage 78, 94–102. http://dx.doi.org/10.1016/j.neuroimage.2013.03.071.

van Gog, T., Kester, L., 2012. A test of the testing effect: acquiring problem-solving skills from worked examples. Cogn. Sci. 36 (8), 1532–1541. http://dx.doi.org/10.1111/cogs.12002.

van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., Verkoeijen, P.P.J.L., 2015. Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. Educ. Psychol. Rev. 27 (2), 265–289. http://dx.doi.org/10.1007/s10648-015-9297-3.

van Gog, T., Sweller, J., 2015. Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. Educ. Psychol. Rev. 27 (2), 247–264. http://dx.doi.org/10.1007/s10648-015-9310-x.

Verkoeijen, P.P.J.L., Bouwmeester, S., Camp, G., 2012. A short-term testing effect in cross-language recognition. Psychol. Sci. 23 (6), 567–571. http://dx.doi.org/10.1177/0956797611435132.

Verkoeijen, P.P.J.L., Tabbers, H.K., Verhage, M.L., 2011. Comparing the effects of testing and restudying on recollection in recognition memory. Exp. Psychol. 58 (6), 490–498. http://dx.doi.org/10.1027/1618-3169/a000117.

Wheeler, M.A., Roediger, H.L., 1992. Disparate effects of repeated testing: reconciling Ballard's (1913) and Bartlett's (1932) results. Psychol. Sci. 3 (4), 240–245. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00036.x.

Whiffen, J.W., Karpicke, J.D., 2017. The role of episodic context in retrieval practice effects. J. Exp. Psychol. Learn. Mem. Cogn. (in press).

Wiklund-Hornqvist, C., Jonsson, B., Nyberg, L., 2014. Strengthening concept learning by repeated testing. Scand. J. Psychol. 55 (1), 10–16. http://dx.doi.org/10.1111/sjop.12093.

Wissman, K.T., Rawson, K.A., Pyc, M.A., 2011. The interim test effect: testing prior material can facilitate the learning of new material. Psychon. Bull. Rev. 18 (6), 1140–1147. http://dx.doi.org/10.3758/s13423-011-0140-7.

Wixted, J.T., Rohrer, D., 1993. Proactive interference and the dynamics of free recall. J. Exp. Psychol. Learn. Mem. Cogn. 19 (5), 1024–1039. http://dx.doi.org/10.1037/0278-7393.19.5.1024.

Yonelinas, A.P., 2002. The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language 46 (3), 441–517. http://dx.doi.org/10.1006/jmla.2002.2864.

Zaromb, F.M., Roediger, H.L., 2010. The testing effect in free recall is associated with enhanced organizational processes. Mem. Cogn. 38 (8), 995–1008. http://dx.doi.org/10.3758/mc.38.8.995.