Review

Cell PRESS

The critical role of retrieval practice in long-term retention

Henry L. Roediger III[1] and Andrew C. Butler[2]

[1] Department of Psychology, Box 1125, Washington University, One Brookings Drive, St. Louis, MO 63130-4899, USA
[2] Psychology & Neuroscience, Duke University, Box 90086, Durham, NC 27708-0086, USA

Learning is usually thought to occur during episodes of studying, whereas retrieval of information on testing simply serves to assess what was learned. We review research that contradicts this traditional view by demonstrating that retrieval practice is actually a powerful mnemonic enhancer, often producing large gains in long-term retention relative to repeated studying. Retrieval practice is often effective even without feedback (i.e. giving the correct answer), but feedback enhances the benefits of testing. In addition, retrieval practice promotes the acquisition of knowledge that can be flexibly retrieved and transferred to different contexts. The power of retrieval practice in consolidating memories has important implications for both the study of memory and its application to educational practice.

Introduction

A curious peculiarity of our memory is that things are impressed better by active than by passive repetition. I mean that in learning (by heart, for example), when we almost know the piece, it pays better to wait and recollect by an effort within, than to look at the book again. If we recover the words the former way, we shall probably know them the next time; if in the latter way, we shall likely need the book once more.

William James [1]

Psychologists have often studied learning by alternating series of study and test trials. In other words, material is presented for study (S) and a test (T) is subsequently given to determine what was learned. After this procedure is repeated over numerous ST trials, performance (e.g. the number of items recalled) is plotted against trials to depict the rate of learning; the outcome is referred to as a learning curve and it is negatively accelerated and is fit by a power function. Thus, most learning occurs on early ST trials, and the amount of learning decreases with additional trials. The critical assumption is that learning occurs during the study phases of the ST ST ST... sequence, and the test phase is simply there to measure what has been learned during previous occasions of study. The test is usually considered a neutral event. For example, researchers in the 1960 s debated whether learning occurs gradually (e.g. through continual strengthening of memory traces) or in an all-or-none fashion, but they focused on study events as the locus of the effects and

ignored the possibility that learning occurred during the retrieval tests [2–5]. Exactly the same assumption is built into our educational systems. Students are thought to learn via lectures, reading, highlighting, study groups, and so on; tests are given in the classroom to measure what has been learned from studying. Again, tests are considered assessments, gauging the knowledge that has been acquired without affecting it in any way.

In this article, we review evidence that turns this conventional wisdom on its head: retrieval practice (as occurs during testing) often produces greater learning and long-term retention than studying. We discuss research that elucidates the conditions under which retrieval practice is most effective, as well as evidence demonstrating that the mnemonic benefits of retrieval practice are transferrable to different contexts. We also describe current theories on the mechanisms underlying the beneficial effects of testing. Finally, we discuss educational implications of this research, arguing that more frequent retrieval practice in the classroom would increase long-term retention and transfer.

The testing effect and repeated retrieval

The finding that retrieval of information from memory produces better retention than restudying the same information for an equivalent amount of time has been termed the testing effect [6]. Although the phenomenon was first reported over 100 years ago [7], research on the testing effect has been sporadic at best until recently (but see Box 1 for some classic studies). In the last 10 years, much research has shown powerful mnemonic benefits of retrieval practice [8–10] . The data in Figure 1 come from a study in which two groups of students retrieved information several times

---

Glossary

Expanding retrieval schedule: testing of retention shortly after learning to make sure encoding is accurate, then waiting longer to retrieve again, then waiting still longer for a third retrieval and so on.
Feedback: providing information after a question. General (right or wrong) feedback is not very helpful if the correct answer is not provided. Correct answer feedback usually produces robust gains on a final criterion measure.
Negative suggestion effect: taking a test that provides subtly wrong answers (e.g. true or false, multiple choice) can lead students to select a wrong answer, believe it is right, and thus learn an error from taking the test.
Retrieval practice: act of calling information to mind rather than rereading it or hearing it. The idea is to produce 'an effort from within' to induce better retention.
Test-enhanced learning: general approach that promotes retrieval practice via testing as a means to improve knowledge.
Testing effect: taking a test usually enhances later performance on the material relative to rereading it or to having no re-exposure at all.
Transfer: ability to generalize learning from one context to another or to use learned information in a new way (e.g. to solve a problem).

Corresponding author: Roediger, H.L. III (roediger@wustl.edu).

during learning and two other groups were treated similarly but only practiced retrieval once [11]. The figure shows performance on a final test given 1 week later. The t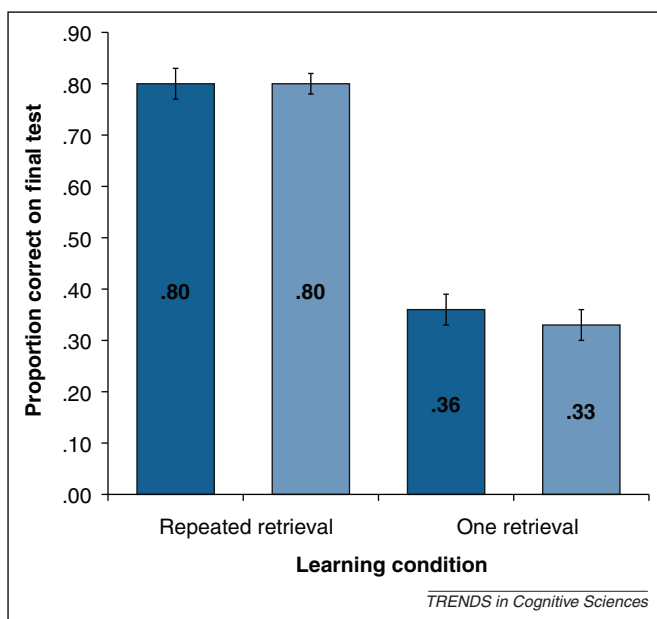wo groups that practiced retrieval (without feedback) during learning (the two left bars) recalled substantially more of the pairs than the other two groups. In addition, the groups represented by the two dark blue bars were permitted to study the material several more times than the groups represented by the light blue bars. Yet, repeated study led to virtually no improvement a week later. Retrieval practice provides much greater long-term retention than does repeated study [11–16].

The finding that retrieval practice increases retention raises two important questions. First, what are the best conditions for retrieval? The sooner retrieval is attempted after a study trial or a correct retrieval, the more likely it is to be successful. Short delays between retrievals might foster errorless retrieval. However, it might be that retrieval of information after a short delay is too much like rote rehearsal, which often produces little or no mnemonic benefit [17]. Second, how many retrievals are needed to maximize long-term retention? Retrieval practice takes time, so if only one or two retrievals is enough, then practice can be terminated [18,19].

The questions just raised are thorny ones and might depend on the type of materials, the characteristics of the learner and other factors (for a discussion see [20]). However, a recent study gives a tentative answer to both questions [21]. Students learned 70 Swahili–English word pairs via repeated practice at retrieving the English word when presented with the associated Swahili word. Both the time between successive retrievals (1 min or 6 min) and the number of successful retrievals (1, 3, 5, 6, 7, 8 or 10) were manipulated during the initial practice phase. Figure 2 shows performance on a final test given after a delay of either 25 min (top two lines) or 1 week (bottom two lines). Regardless of the timing of the final test, retrieval practice with 6-min intervening intervals (red lines) led to better retention relative to retrieval practice with 1-min intervening intervals (blue lines). With respect to the number of

Figure 1. Recall after a week for Swahili–English word pairs (mashua–boat) learned with retrieval practice (left bars) or with only a single recall (right bars). Retrieval practice doubled recall on the final test when students were given the Swahili word and asked to recall the English word. The dark blue bars indicate groups to which many more study trials were given than to the groups represented by light blue bars. Repetition of studying had virtually no effect on recall a week later, unlike repeated retrieval. Error bars represent standard errors of the mean. Figure adapted from [11].

Figure 2. Recall after 25 min (top two lines) or 1 week (bottom two lines) after varying numbers of correct recalls in an earlier phase of the experiment. When 6 min occurred between retrievals (red lines), performance was better than when only 1 min occurred between tests (blue lines). When only a short interval occurred between retrievals, even recalling the pair ten times failed to improve retention a week later. Figure adapted from [21].

successful retrievals during initial learning, final test performance generally increased from one to five or seven prior retrievals and then leveled off, so five to seven retrievals seem to be optimal in this paradigm. However, this pattern of performance depended on the time between successive retrievals during initial practice. After a week, only retrieval practice with longer intervening intervals had any effect on performance – practice that occurred every minute produced floor-level performance, no matter how many times the item was successfully retrieved.

Retrieval practice can be a potent memory enhancer, but clearly the conditions of retrieval matter. When retrieval occurs under relatively easy (1-min interval) conditions, even ten retrievals might produce little benefit for long-term retention. By contrast, under different conditions, many other studies have shown that even a single test can boost retention [22,23] and that these benefits persist over long delays [14,24]. Still, repeated retrievals usually benefit later retention relative to a single retrieval [14,21,25,26].

Expanding retrieval schedules

The data in Figure 2 might be considered surprising in some quarters. For example, researchers who perform behavior analysis [27] or memory remediation among neuropsychological patient populations [28] believe that retrieval attempts should be arranged so that they do not produce errors (errorless retrieval is the watchword in these efforts). The fear is that if an error is produced than it will be learned, making learning of the correct responses more difficult. However, the data in Figure 2 point to a paradox: if retrieval occurs under 'easy' conditions in which errors are less likely to be made, the impact of such retrievals on long-term retention might be undermined. Thus, a practical question is whether a strategy exists for retrieval practice that precludes making errors and at the same time permits the type of difficult retrievals that produce better long-term retention.

One possible strategy is the expanding schedule of retrieval, which was first proposed by Landauer and Bjork [29]. In this method, a first retrieval attempt occurs shortly after initial learning and subsequent retrieval attempts are staggered so that each successive retrieval occurs after an increasingly long interval. For example, when learning someone's name, retrieval of the name would occur shortly after meeting the person (say, 1 min) to be sure it is encoded, then after a slightly longer interval (perhaps 4 min), and then after a still longer interval (8 min) before retrieving it a third time, and so on. The idea is to gradually shape long-term retention of the information just as learning can be shaped by reinforcement of successive approximations of the desired behavior [30].

In their influential paper, Landauer and Bjork predicted that expanding retrieval schedules would produce better performance than equal-interval schedules (in which the intervals between retrieval attempts remain constant) or massed schedules (repeated retrieval with no intervening interval) [29]. Indeed, findings from their experiments showed a benefit of an expanding schedule relative to an equal-interval schedule on a final test given after a relatively short retention interval of 30 min. Furthermore,

both the expanding and equal-interval schedules produced better final retention than did a massed schedule of practice, even though the massed tests provided nearly errorless retrieval. Thus, research comparing different schedules of practice provides additional evidence that repeated retrieval of information immediately after study, even though errorless, produces poor retention [31–33].

Returning to the issue of whether expanding or equal-interval schedules of practice lead to better retention, the answer seems to depend in part on the retention interval. When the final test is given shortly after the learning phase, expanding retrieval seems to be best. However, when long-term retention is measured (i.e. a delay of a day or longer), then prior practice on an equal-interval schedule seems to promote better performance [34,35]. The reason for this flip in performance from immediate to delayed tests might be due to the timing of the initial test: the first test is given almost immediately in an expanding schedule, whereas it is given after a longer delay in an equal-interval schedule. Thus, the equal-interval schedule requires greater retrieval effort on the first test, which should produce better long-term retention. The general conclusion is that the best retrieval schedules are those that involve wide spacing of retrieval attempts, as shown in Figure 2 [21], even if some errors are made [36,37]. To date, evidence shows that expanding retrieval provides better retention after short delays, but equal interval retrieval produces better retention after long delays. However, expanding schedules may show a benefit in future research with expansion that unfolds over days and weeks rather than over seconds (as used in past research).

Feedback enhances the testing effect

Although retrieval practice promotes superior long-term retention in the absence of feedback (Figure 1), providing the correct answer after a retrieval attempt increases the mnemonic benefits of testing [38,39]. Feedback that includes the correct answer increases learning because it enables test-takers to correct errors [40] and to maintain correct responses [41]. The critical mechanism in learning from tests is successful retrieval; however, if test-takers do not retrieve the correct response and have no recourse to learn it, then the benefits of testing can sometimes be limited or absent altogether [42]. Thus, providing feedback after a retrieval attempt, regardless of whether the attempt is successful or unsuccessful, helps to ensure that retrieval will be successful in the future [41].

The need for feedback is critical after any type of test, but it is particularly important for recognition tests (e.g. multiple choice, true/false, etc.) because test-takers are exposed to incorrect information. For example, on multiple-choice tests, students must identify the correct answer from a number of possible alternative answers (i.e. lures), most of which are plausible but incorrect. The danger is that because students learn from tests, taking a multiple-choice test might cause them to learn incorrect information and believe that it is true. Indeed, recent research has shown that when students select a lure in a multiple-choice test, they often reproduce that incorrect information in a later test [8,43,44]. This outcome even occurs on the SAT test that hundreds of thousands of high school students

take every year [45]. Although the potential for negative effects from multiple-choice tests is a real problem, the good news is that there is a simple solution: provide students with feedback. If feedback is provided after a multiple-choice test, the negative effects are completely nullified [46]. Thus, whereas feedback is helpful for all types of tests, it is especially important for multiple-choice and other recognition tests that can lead students to learn incorrect information.

Another critical question is the timing of feedback. Conventional wisdom and studies in behavioral psychology indicate that providing feedback immediately after a test is best [27,47]. However, experimental results show that delayed feedback might be even more powerful. In one study, students read passages and then either took or did not take a multiple-choice test [16]. For students who took the test, one group received correct answer feedback immediately after making a response (immediate feedback) and the other group received the correct answers for all questions after the entire test (delayed feedback). One week after the initial learning session, students took a final test in which they had to produce a response to the question that had formed the stem of the multiple-choice item (i.e. they had to produce the answer rather than selecting one from among several alternatives). The final test consisted of the same questions from the initial multiple-choice test and comparable questions that had not been tested.

Figure 3 shows the results for the final test. Taking an initial test (even without feedback) tripled final recall relative to only studying the material. When correct answer feedback was given immediately after each question in the initial test, performance increased another 10%. However, feedback given after the entire test boosted final performance even more. The finding that delayed feedback led to better retention than immediate feedback undermines the conventional idea that feedback must be given

immediately to be effective. Although giving the answers to questions soon after a test is still relatively immediate feedback, the superiority of delayed feedback has been replicated numerous times with longer delays [48–51]. The benefits of delayed feedback might represent a type of spacing effect: the phenomenon whereby two presentations of material given with spacing between them generally leads to better retention than massed (back-to-back) presentations [52–55].

Retrieval practice enhances transfer of learning

Are the mnemonic benefits of testing limited to the learning of a specific response? One criticism that could be leveled at research on the testing effect is that retrieval practice merely teaches people to produce a fixed response when given a particular retrieval cue, so the procedure merely amounts to drill and practice of a particular response. Thus, a key question is whether testing also promotes transfer of knowledge; that is, can the knowledge gained through testing be flexibly used to construct new responses and answer different questions? Transfer of learning is of critical interest for both theories of memory and educational policy [56]. Researchers have recently begun to explore whether retrieval practice can promote transfer of learning in different contexts [57–59].

For example, Butler [60] investigated whether repeated testing produces better transfer than repeated studying in a series of experiments. In one of the experiments, students studied six prose passages, each of which contained several critical concepts (among other information). A concept was operationally defined as information that had to be extracted from multiple sentences. Next, the students repeatedly restudied two of the passages, repeatedly restudied isolated sentences that contained the critical concepts from another two passages, and repeatedly took a test on the critical concepts for another two passages. After each test
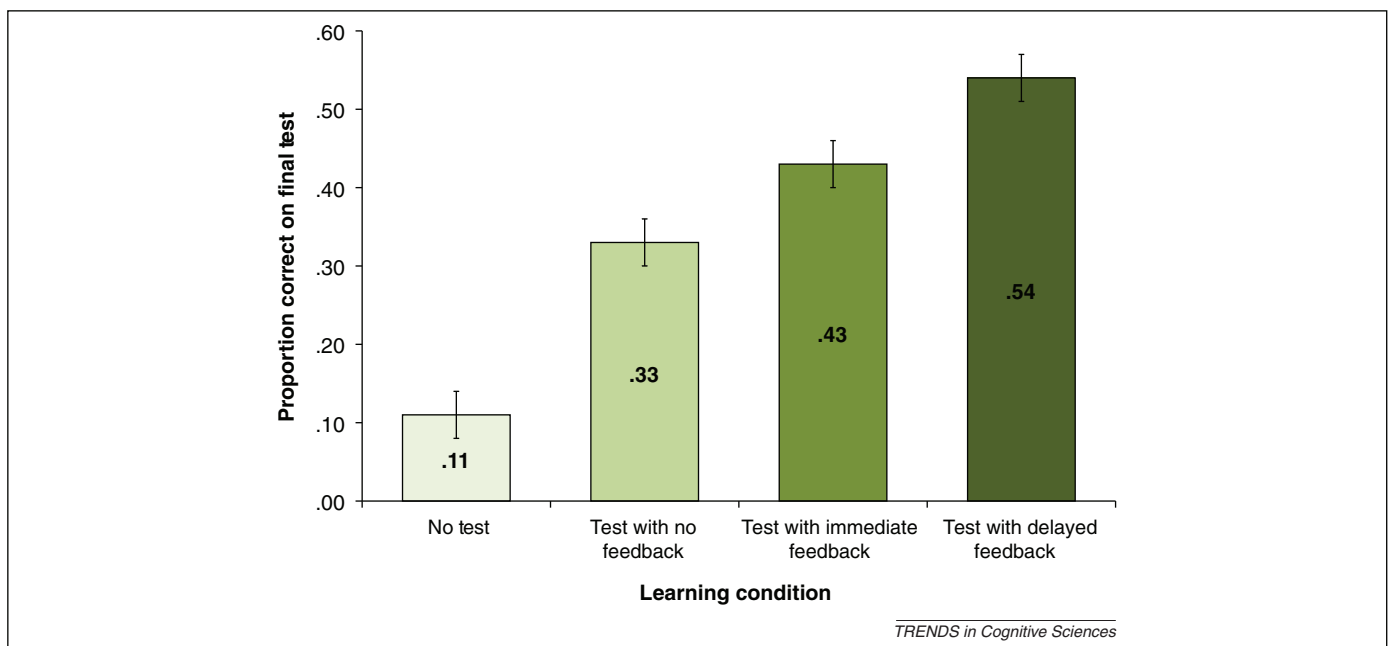
Figure 3. Proportion of correct responses on the final cued recall test as a function of initial learning condition. All conditions involving an initial test led to greater final recall than in the No Test condition, but feedback after the initial test led to greater final recall. In addition, delayed feedback (given on each item after the test) led to better recall than did immediate feedback (given after each question was answered). Error bars represent 95% confidence intervals. The figure represents data in Table 2 from [46].

Box 2. Sample materials from Butler [60]

The passages used in the study covered a range of topics. The questions below are samples from a passage about bats.

Initial test
Question: Some bats use echolocation to navigate the environment and locate prey. How does echolocation help bats to determine the distance and size of objects?

Answer: Bats emit high-pitched sound waves and listen to the echoes. The distance of an object is determined by the time it takes for the echo to return. The size of the object is calculated by the intensity of the echo: a smaller object will reflect less of the sound wave, and thus produce a less intense echo.

Final transfer test
Question: An insect is moving towards a bat. Using the process of echolocation, how does the bat determine that the insect is moving towards it (i.e. rather than away from it)?

Answer: The bat can tell the direction that an object is moving by calculating whether the time it takes for an echo to return changes from echo to echo. If the insect is moving towards the bat, the time it takes the echo to return will get steadily shorter.

---

question, students received feedback that was essentially the same information as that presented in the condition with the restudied isolated sentences. Thus, the key difference between the restudy isolated sentences condition and the repeated testing condition was that students attempted to retrieve the information in the latter condition before getting it to restudy. One week later, students took a final test that required application of each critical concept from the passages to a new inferential question from the same knowledge domain. Examples of materials are shown in Box 2.

Figure 4 shows results for the final test. Interestingly, there was virtually no difference between the two repeated
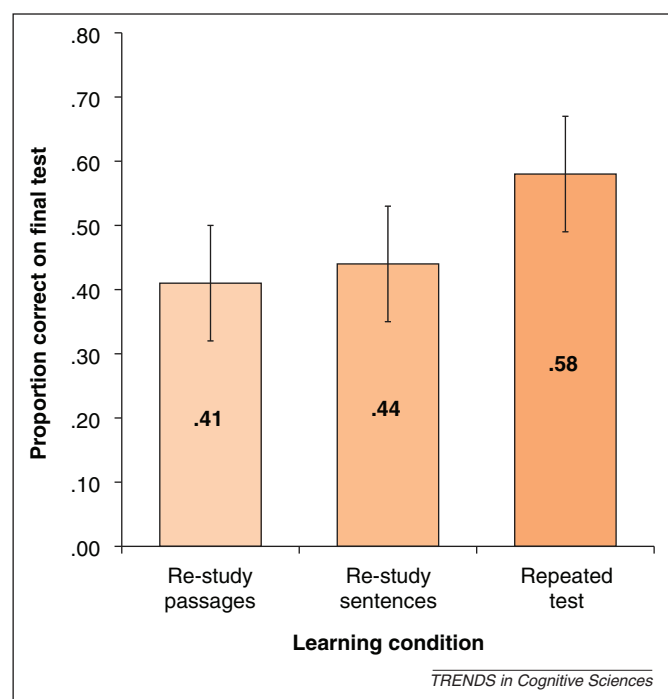

Figure 4. Proportion of correct responses on the final cued recall test as a function of initial learning condition. The retrieval practice (testing) conditions led to greater transfer relative to repeated restudying of whole passages or restudying of just the sentences containing the critical concepts. Error bars represent 95% confidence intervals. Figure adapted from [60].

study conditions even though studying the isolated sentences ostensibly allowed for more time to learn the critical concepts than studying the entire passage. This result fits well with the findings of other studies demonstrating that restudying provides limited benefits for retention (Figure 1) [61]. More importantly, repeated testing led to significantly better transfer than either repeated studying of the passages or repeated studying of the isolated sentences. This finding indicates that the mnemonic benefits of testing extend well beyond the retention of a specific response. In fact, a subsequent experiment in the same series showed that repeated testing produced better transfer relative to repeated studying on new inferential questions about different knowledge domains (e.g. applying knowledge about echolocation in bats to sonar in submarines), a situation that constitutes far transfer according to one definition [56].

Theories of the retrieval practice effects
Researchers have intensively studied the effects of retrieval practice and today we know much about conditions that produce the effect. However, theoretical understanding – or even proper theories of the effect – has lagged behind. One idea sometimes invoked to explain retrieval practice (testing) effects is that such practice simply permits re-exposure to material and causes overlearning of the set of material that can be retrieved [62,63]. Many experiments have discredited this hypothesis by showing that equating the number of study events to test (retrieval) events does not eliminate the effect [6,64,65]. The data in Figure 2 also show that this idea must be wrong, because with number of retrievals equated at various levels, some conditions produced huge retrieval practice effects and others none at all.

In general, theoretical explanations for retrieval practice (testing) effects have focused on how the act of retrieval affects memory. One idea is that retrieval of information from memory leads to elaboration of the memory trace and/or the creation of additional retrieval routes, which makes it more likely that the information will be successfully retrieved again in the future [22,66,67]. A related idea invokes the notion of retrieval effort to explain the positive effects of retrieval practice [21,68]. Retrieval effort can be thought of as an index of the amount of reprocessing of the memory trace that occurs during retrieval: the more effort involved in retrieving the memory, the more extensive is the reprocessing (which presumably involves elaboration). As discussed above, retrieval practice that occurs under conditions in which information can be easily accessed (e.g. from short-term or working memory) leads to little or no benefit for long-term retention (Figure 2). Yet another explanation relies on the concept of transfer-appropriate processing [69,70], which holds that memory performance is enhanced to the extent that the cognitive processes during learning match those required during retrieval. The processes engaged by taking an initial test provide a better match with final test than the processes involved in studying the material.

The new theory of disuse of Bjork and Bjork incorporates these ideas to provide a more formal explanation of retrieval practice effects [71]. The theory distinguishes between storage strength (relative permanence of the

memory trace) and retrieval strength (momentary accessibility of a trace). For example, if a weak trace (in terms of storage strength) has recently been retrieved, its retrieval strength will be great for some time afterward. The theory proposes that positive effects of retrieval on storage strength are inversely related to retrieval strength; the greater the retrieval strength, the less is the effect of retrieval on storage strength. This idea would account for the fact that repeated retrieval just after study has little effect and other data such as those in Figure 2.

The theories above and others [72] are psychological ones at an abstract level of description. Mechanistic accounts of testing in neuroscientific terms await development. However, we can point to some promising leads. The concept of reconsolidation – the idea that retrieval of a memory places it into a labile state in which the trace can be enhanced or disrupted – has become a topic of considerable interest in neuroscience in the past 10 years [73,74]. The molecular cascade involved in reconsolidation [75,76] will doubtless be involved in explaining the mnemonic benefits of retrieval practice. Interaction between the hippocampus and dopaminergic neurons in the ventral tegmental area (VTA) might provide another piece of the puzzle [77]. When the hippocampus detects information that is relatively unfamiliar, the novelty signal causes firing of dopaminergic cells, which enhances long-term potentiation and thus learning. Retrieval practice might activate the hippocampal–VTA feedback loop, thereby strengthening connections between the neurons that form the memory trace for the retrieved information. However, this process would only occur when the information is relatively unfamiliar (perhaps having low retrieval strength, in terms used above [71]). These ideas are clearly speculative, but might point the way to a more mechanistic account of retrieval practice effects.

Educational implications

Retrieval practice produces greater long-term retention than studying alone. This finding suggests that testing, which is commonly conceptualized as an assessment tool, can be used as a learning tool as well [78]. In particular, practicing retrieval is beneficial when it requires effortful processing (e.g. production rather than recognition tests), it occurs multiple times with relatively long intervals between retrieval attempts, and it is followed by feedback after each attempt. Under these conditions, tests provide a highly effective means of learning. Educators sometimes decry this approach of what we have called test-enhanced learning [6,9] as involving nothing but drill and practice in which students engage in rote rehearsal. However, when used correctly, retrieval practice techniques help to foster deeper learning and understanding so that knowledge can be flexibly retrieved and transferred to new situations [57–60].

Studies on retrieval practice conducted in educational settings have shown that frequent testing produces substantial benefits to long-term retention [79]. For example, research has demonstrated that retrieval practice improves scores in college courses in biological psychology and statistics [80,81], as well as advanced medical education [82]. In addition, experiments in middle-school history, social studies and science classrooms have shown great

improvement in children's knowledge derived from repeated quizzing on delayed tests [83–85]. Importantly, the tests used to measure long-term retention in some of these studies were the actual tests being given to the class for assessment purposes, not ones made up for the sake of an experiment.

Testing at the university level provides an indirect benefit that complements the direct benefit that is discussed here. Many university courses require only one or two semester tests and a final exam, a practice that leads to the near universal phenomenon of students concentrating their study attempts just before the exams and not keeping up with the course [86,87]. Frequent quizzing (say, on a weekly or even a daily basis) forces students to stay current with the course by studying more regularly. Classroom studies have shown that students who received daily quizzes performed better than those who did not [81,88]. Importantly, survey questions given at the end of the semester revealed that the students who were frequently quizzed felt they had learned more and reported greater satisfaction with the course, despite (or perhaps because of) the greater effort they exerted [81,88]. In addition, the mnemonic benefits of testing extend beyond the specific information that is tested: retrieval practice can increase retention of related, but non-tested material as well [89–91]. Of course, retrieval practice need not occur only through quizzing or testing in the classroom. Retrieval practice can be implemented in many different ways, including self-testing (e.g. using flash cards, chapter-ending questions, or other methods).

Concluding remarks

The finding that retrieval practice yields substantial mnemonic benefits validates the quote from William James [1] at the outset: Students' 'active repetition' via attempts to 'recollect by an effort from within' provides a much greater boost to retention than does 'passive repetition' from an outside source. The research we reviewed makes five points. First, retrieval practice often produces superior long-term retention relative to studying for an equivalent amount of time. Second, repeated testing is better than taking a single test. Third, testing with feedback leads to greater benefits than does testing without feedback, but even the latter procedure can be surprisingly effective. Fourth, to place a caveat on the first three claims, testing under conditions that make retrieval easy (e.g. learning a face–name pair and being tested on it several times immediately) often has surprisingly little effect; some lag between study and test is required for retrieval practice to provide a benefit. Fifth, the mnemonic benefits of retrieval practice are not limited to the learning of a specific response, but rather produce knowledge that can be transferred to different contexts. Integration of retrieval practice into educational practices has the potential to boost performance in schools. Further research is required, however, to understand the mechanisms that give rise to the beneficial effects of retrieval practice.

References

1 James, W. (1890) *The Principles of Psychology*, Holt

2 Estes, W.K. (1960) Learning theory and the new 'mental chemistry'. *Psychol. Rev.* 67, 207–223

3 Postman, L. (1963) One-trial learning. In *Verbal Behavior and Learning: Problems and Processes* (Cofer, C.N. and Musgrave, B.S., eds), pp. 295–335, McGraw-Hill

4 Rock, I. (1957) The role of repetition in associative learning. *Am. J. Psychol.* 70, 186–193

5 Underwood, B.J. and Keppel, G. (1962) One-trial learning? *J. Verb. Learn. Verb. Behav.* 1, 1–13

6 Roediger, H.L., III and Karpicke, J.D. (2006) The power of testing memory: basic research and implications for educational practice. *Persp. Psychol. Sci.* 1, 181–210

7 Abbott, E.E. (1909) On the analysis of the factors of recall in the learning process. *Psychol. Monogr.* 11, 159–177

8 Marsh, E.J. *et al.* (2007) The memorial consequences of multiple-choice testing. *Psychonom. Bull. Rev.* 14, 194–199

9 McDaniel, M.A. *et al.* (2007) Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonom. Bull. Rev.* 14, 200–206

10 Pashler, H. *et al.* (2007) Enhancing learning and retarding forgetting: choices and consequences. *Psychonom. Bull. Rev.* 14, 187–193

11 Karpicke, J.D. and Roediger, H.L., III (2008) The critical importance of retrieval for learning. *Science* 15, 966–968

12 Carpenter, S.K. *et al.* (2008) The effects of tests on learning and forgetting. *Mem. Cogn.* 36, 438–448

13 Kuo, T. and Hirshman, E. (1996) Investigations of the testing effect. *Am. J. Psychol.* 109, 451–464

14 Roediger, H.L., III and Karpicke, J.D. (2006) Test-enhanced learning: taking memory tests improves long-term retention. *Psychol. Sci.* 17, 249–255

15 Toppino, T.C. and Cohen, M.S. (2009) The testing effect and the retention interval: questions and answers. *Exp. Psychol.* 56, 252–257

16 Wheeler, M.A. *et al.* (2003) Different rates of forgetting following study versus test trials. *Memory* 11, 571–580

17 Craik, F.I.M. and Watkins, M.J. (1973) The role of rehearsal in short-term memory. *J. Verb. Learn. Verb. Behav.* 12, 599–607

18 Pyc, M.A. and Rawson, K.A. (2007) Examining the efficiency of schedules of distributed retrieval practice. *Mem. Cogn.* 35, 1917–1927

19 Kornell, N. and Bjork, R.A. (2008) Optimising self-regulated study: the benefits – and costs – of dropping flashcards. *Memory* 16, 125–136

20 McDaniel, M.A. and Butler, A.C. (2010) A contextual framework for understanding when difficulties are desirable. In *Successful Remembering and Successful Forgetting: Essays in Honor of Robert A. Bjork* (Benjamin, A.S., ed.), pp. 175–199, Psychology Press

21 Pyc, M.A. and 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, 437–447

22 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, 1563–1569

23 Carpenter, S.K. and DeLosh, E.L. (2006) Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. *Mem. Cogn.* 34, 268–276

24 Butler, A.C. and Roediger, H.L., III (2007) Testing improves long-term retention in a simulated classroom setting. *Eur. J. Cogn. Psychol.* 19, 514–527

25 Hogan, R.M. and Kintsch, W. (1971) Differential effects of study and test trials on long-term recognition and recall. *J. Verb. Learn. Verb. Behav.* 10, 562–567

26 Wheeler, M.A. and Roediger, H.L., III (1992) Disparate effects of repeated testing: reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychol. Sci.* 3, 240–245

27 Skinner, B.F. (1954) The science of learning and the art of teaching. *Harv. Educ. Rev.* 24, 86–97

28 Baddeley, A.D. and Wilson, B.A. (1994) When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia* 32, 53–68

29 Landauer, T.K. and Bjork, R.A. (1978) Optimum rehearsal patterns and name learning. In *Practical Aspects of Memory* (Gruneberg, M.M. *et al.*, eds), pp. 625–632, Academic Press

30 Skinner, B.F. (1953) *Science and Human Behavior*, Macmillan

31 Cull, W.L. (2000) Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Appl. Cogn. Psychol.* 14, 215–235

32 Cull, W.L. *et al.* (1996) Expanding understanding of the expanding-pattern-of-retrieval mnemonic: toward confidence in applicability. *J. Exp. Psychol. Appl.* 2, 365–378

33 Balota, D.A. *et al.* (2007) Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In *The Foundations of Remembering: Essays in Honor of Henry L. Roediger, III* (Nairne, J.S., ed.), pp. 83–106, Psychology Press

34 Karpicke, J.D. and Roediger, H.L., III (2007) Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 704–719

35 Logan, J.M. and Balota, D.A. (2008) Expanded vs. equal interval spaced retrieval practice: exploration of schedule of spacing and retention interval in younger and older adults. *Aging Neuropsychol. Cogn.* 15, 257–280

36 Roediger, H.L., III and Karpicke, J.D. (2010) Intricacies of spaced retrieval: a resolution. In *Successful Remembering and Successful Forgetting: Essays in Honor of Robert A. Bjork* (Benjamin, A.S., ed.), Psychology Press, pp. 23–47

37 Pashler, H. *et al.* (2003) Is temporal spacing of tests helpful even when it inflates error rates? *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1051–1057

38 Bangert-Drowns, R.L. *et al.* (1991) The instructional effect of feedback in test-like events. *Rev. Educ. Res.* 61, 213–238

39 Kulhavy, R.W. and Stock, W.A. (1989) Feedback in written instruction: the place of response certitude. *Educ. Psychol. Rev.* 1, 279–308

40 Pashler, H. *et al.* (2005) When does feedback facilitate learning of words? *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 3–8

41 Butler, A.C. *et al.* (2008) Correcting a meta-cognitive error: feedback enhances retention of low confidence correct responses. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 918–928

42 Kang, S.H.K. *et al.* (2007) Test format and corrective feedback modulate the effect of testing on memory retention. *Eur. J. Cogn. Psychol.* 19, 528–558

43 Butler, A.C. *et al.* (2006) When additional multiple-choice lures aid versus hinder later memory. *Appl. Cogn. Psychol.* 20, 941–956

44 Roediger, H.L., III and Marsh, E.J. (2005) The positive and negative consequences of multiple-choice testing. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1155–1159

45 Marsh, E.J. *et al.* (2009) Memorial consequences of answering SAT II questions. *J. Exp. Psychol. Appl.* 15, 1–11

46 Butler, A.C. and Roediger, H.L., III (2008) Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem. Cogn.* 36, 604–616

47 Kulik, J.A. and Kulik, C.C. (1988) Timing of feedback and verbal learning. *Rev. Educ. Res.* 58, 79–97

48 Butler, A.C. *et al.* (2007) The effect of type and timing of feedback on learning from multiple-choice tests. *J. Exp. Psychol. Appl.* 13, 273–281

49 Kulhavy, R.W. and Anderson, R.C. (1972) Delay-retention effect with multiple-choice tests. *J. Educ. Psychol.* 63, 505–512

50 Metcalfe, J. *et al.* (2009) Delayed versus immediate feedback in children's and adults' vocabulary learning. *Mem. Cogn.* 37, 1077–1087

51 Smith, T.A. and Kimball, D.R. (2010) Learning from feedback: spacing and the delay-retention effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 80–95

52 Cepeda, N.J. *et al.* (2006) Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132, 354–380

53 Cepeda, N.J. *et al.* (2008) Spacing effect in learning: a temporal ridgeline of optimal retention. *Psychol. Sci.* 19, 1095–1102

54 Melton, A.W. (1970) The situation with respect to the spacing of repetitions and memory. *J. Verb. Learn. Verb. Behav.* 9, 596–606

55 Madigan, S.A. (1969) Intraserial repetition and coding processes in free recall. *J. Verb. Learn. Verb. Behav.* 8, 828–835

56 Barnett, S.M. and Ceci, S.J. (2002) When and where do we apply what we learn? A taxonomy for far transfer. *Psychol. Bull.* 128, 612–637

57 Johnson, C.I. and Mayer, R.E. (2009) A testing effect with multimedia learning. *J. Educ. Psychol.* 101, 621–629

58 McDaniel, M.A. *et al.* (2009) The read–recite–review study strategy: effective and portable. *Psychol. Sci.* 20, 516–522

59 Rohrer, D. *et al.* (2010) Tests enhance the transfer of learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 233–239

60 Butler, A.C. (2010) Repeated testing produces superior transfer of learning relative to repeated studying. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1118–1133

61 Callender, A.A. and McDaniel, M.A. (2009) The limited benefits of rereading educational texts. *Contemp. Educ. Psychol.* 34, 30–41

62 Slamecka, N.J. and Katsaiti, L.T. (1988) Normal forgetting of verbal lists as a function of prior testing. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 716–727

63 Thompson, C.P. *et al.* (1978) How recall facilitates subsequent recall: a reappraisal. *J. Exp. Psychol. Hum. Learn. Mem.* 4, 210–221

64 Glover, J.A. (1989) The "testing" phenomenon: not gone but nearly forgotten. *J. Educ. Psychol.* 81, 392–399

65 Carrier, M. and Pashler, H. (1992) The influence of retrieval on retention. *Mem. Cogn.* 20, 633–642

66 Bjork, R.A. (1975) Retrieval as a memory modifier: an interpretation of negative recency and related phenomena. In *Information Processing and Cognition* (Solso, R.L., ed.), pp. 123–144, Wiley

67 McDaniel, M.A. and Masson, M.E.J. (1985) Altering memory representations through retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 371–385

68 Gardiner, J.M. *et al.* (1973) Retrieval difficulty and subsequent recall. *Mem. Cogn.* 1, 213–216

69 Morris, C.D. *et al.* (1977) Levels of processing versus transfer-appropriate processing. *J. Verb. Learn. Verb. Behav.* 16, 519–533

70 Roediger, H.L., III *et al.* (2002) Processing approaches to cognition: the impetus from the levels of processing framework. *Memory* 10, 319–332

71 Bjork, R.A. and Bjork, E.L. (1992) A new theory of disuse and an old theory of stimulus fluctuation. In *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes* (Vol. 2) (Healy, A. *et al.*, eds), pp. 35–67, Erlbaum.

72 Pavlik, P.I., Jr (2007) Understanding and applying the dynamics of test practice and study practice. *Instruct. Sci.* 35, 407–441

73 Dudai, Y. (2004) The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* 55, 51–86

74 Sara, S.J. (2000) Retrieval and reconsolidation: toward a neurobiology of remembering. *Learn. Mem.* 7, 73–84

75 Lee, J.L. *et al.* (2004) Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science* 304, 839–843

76 Nader, K. *et al.* (2000) Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* 406, 722–726

77 Lisman, J.E. and Grace, A.A. (2005) The hippocampal–VTA loop: controlling the entry of information into long-term memory. *Neuron* 46, 703–713

78 Dempster, F.N. (1992) Using tests to promote learning: a neglected classroom resource. *J. Res. Dev. Educ.* 25, 213–217

79 Bangert-Drowns, R.L. *et al.* (1991) The instructional effect of feedback in test-like events. *Rev. Educ. Res.* 61, 213–238

80 McDaniel, M.A. *et al.* (2007) Testing the testing effect in the classroom. *Eur. J. Cogn. Psychol.* 19, 494–513

81 Lyle, K.B. and Crawford, N.A. Retrieving essential material at the end of lectures improves performance on statistics exams. *Teach. Psychol.* in press.

82 Larsen, D.P. *et al.* (2009) Repeated testing improves long-term retention relative to repeated study: a randomized, controlled trial. *Med. Educ.* 43, 1174–1181

83 Carpenter, S.K. *et al.* (2009) Using tests to enhance 8th grade students' retention of U.S. history facts. *Appl. Cogn. Psychol.* 23, 760–771

84 Roediger, H.L., III et al. (submitted). Test-enhanced learning in the classroom: Long-term improvements from quizzing.

85 McDaniel, M.A. *et al*. Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. *J. Educ. Psychol*. in press.

86 Mawhinney, V.T. *et al.* (1971) A comparison of students studying-behavior produced by daily, weekly, and three-week testing schedules. *J. Appl. Behav. Anal.* 4, 257–264

87 Michael, J. (1991) A behavioral perspective on college teaching. *Behav. Anal.* 14, 229–239

88 Leeming, F.C. (2002) The exam-a-day procedure improves performance in psychology classes. *Teach. Psychol.* 29, 210–212

89 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, 153–170

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

91 Chan, J.C.K. *et al.* (2006) Retrieval induced facilitation: initially nontested material can benefit from prior testing. *J. Exp. Psychol. Gen.* 135, 533–571

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

93 Jones, H.E. (1923-1924) The effects of examination on the performance of learning. *Arch. Psychol.* 10, 1–70

94 Spitzer, H.F. (1939) Studies in retention. *J. Educ. Psychol.* 30, 641–656

95 Tulving, E. (1967) The effects of presentation and recall of material in free-recall learning. *J. Verb. Learn. Verb. Behav.* 6, 175–184