

Planning Education for Long-Term Retention: The Cognitive Science and Implementation of Retrieval Practice

Douglas P. Larsen, MD, MEd¹

¹ Department of Neurology, Washington University School of Medicine, Washington University in St. Louis, St. Louis, Missouri

Semin Neurol 2018;38:449–456.

Address for correspondence Douglas P. Larsen, MD, MEd, Department of Neurology, Washington University School of Medicine, Washington University in St. Louis, 660 S. Euclid Avenue, Campus Box 8111, St. Louis, MO 63110 (e-mail: larsend@wustl.edu).

Abstract

Educational systems are rarely designed for long-term retention of information. Strong evidence has emerged from cognitive psychology and applied education studies that repeated retrieval of information significantly improves retention compared to repeated studying. This effect likely emerges from the processes of memory consolidation and reconsolidation. Consolidation and reconsolidation are the means by which memories are organized into associational networks or schemas that are created and recreated as memories are formed and recalled. As educators implement retrieval practice, they should consider how various test formats lead to different degrees of schema activation. Repeated acts of retrieval provide opportunities for schemas to be updated and strengthened. Spacing of retrieval allows more consolidated schemas to be reactivated. Feedback provides metacognitive monitoring to ensure retrieval accuracy and can lead to shifts from ineffective to effective retrieval strategies. By using the principles of retrieval practice, educators can improve the likelihood that learners will retain information for longer periods of time.

Keywords

- ▶ memory
- ▶ learning
- ▶ retrieval practice
- ▶ test-enhanced learning
- ▶ schema
- ▶ medical education

An Education Problem

A neurology ward team stands in the hallway discussing a patient with a rare disease that they have just seen. The attending, an expert in the condition, turns to the third-year medical student on the team and begins to “pimp” the student by asking various questions about the disease. Painfully, the student clearly has no idea of the answers. In dismay, the attending asks, “What? Weren’t you at my lecture during second year on this topic?!”

While the episode above may seem comical, it is repeated in various forms throughout our medical education system on a frequent basis. The assumption that teaching results in durable learning is foundational to many of our educational activities: the lectures given in medical school and continuing medical education (CME) courses, the didactic sessions

used to standardize the curriculum in clinical rotations and residency programs, and even the bedside teaching that often accompanies rounds in the hospital. While forgetting is a known and natural physiological phenomenon, our educational systems often do not account for it. Rather we, like the attending above, assume that once a learner has acquired new knowledge that he or she will always possess that information. At the end of courses, clerkships, and even at the end of a given period of training such as the end of medical school or residency, we give tests to measure the accumulated knowledge to ensure that it is adequate and to verify that the learner has achieved a minimum level of competency. We then expect that they will retain this information for the remainder of their careers. Learners have adapted to the incentives created by these systems and will typically engage in intense periods of study just prior to tests to demonstrate the highest possible amount of “learning.” Learners often anecdotally observe that this

Issue Theme Neurology Education;
Guest Editor, Jeremy J. Moeller, MD,
MSc, FRCPC

Copyright © 2018 by Thieme Medical
Publishers, Inc., 333 Seventh Avenue,
New York, NY 10001, USA.
Tel: +1(212) 584-4662.

DOI <https://doi.org/10.1055/s-0038-1666983>.
ISSN 0271-8235.

strategy leads to rapid forgetting—referring to a “binge/purge cycle” in their schooling. Long-term retention of information is rarely a planned element in the educational system.

Despite these challenges, research from cognitive science has emerged that can guide educators to use tools and create systems that do promote long-term retention of information. One of these tools is *retrieval practice*. The cognitive science research regarding retrieval practice has been extensively reviewed elsewhere.^{1,2} However, this review will focus on a synthesis of the educational practices that can be recommended based on this body of research. While Larsen and Butler have also reviewed these principles previously,³ the current review will discuss these principles in light of possible neural mechanisms and consider the challenges and opportunities of integrating retrieval practice into curricula.

What is Retrieval Practice?

As described above, tests are typically thought of as measurement devices to evaluate the extent of learning without necessarily changing the memory of that knowledge. However, extensive research has emerged that retrieval of information (whether in tests or through other experiences) actually leads to more enduring memory for the material when compared with restudying.^{1,2} This phenomenon is referred to as the *direct testing effect*.¹ The act of retrieving information to improve long-term retention is known as *retrieval practice*.² It is important to specify that the direct testing effect referred to here arises from the act of retrieval and its associated cognitive processes not from improved studying or increased motivation to study in anticipation of taking a test (indirect testing effects). While tests may have such positive behavioral effects, this review will focus on the direct cognitive effects that arise from retrieval.

Research on the direct testing effect has existed for more than 100 years.^{4,5} However, broad interest in this finding and extension into educational practice is a more recent phenomenon. One of the experiments that launched this wave of new research was performed with college students using science texts.⁶ Students were randomized to one of three conditions: studying the text by re-reading it four times, re-reading it three times and then taking a free recall test in which they were asked to simply write down everything that they could remember from the text, or reading the text once and then taking three back-to-back free recall tests without any feedback. Students were then either given a final recall test 5 minutes later or a final test 1 week later. On the 5-minute test, students who had learned the material through re-reading remembered the most, followed by those who had read the text three times and taken one test, while those who had read the text only once but had taken three tests remembered the least. However, on the final test 1 week later, an inverse pattern was observed. Those who had taken three tests remembered the most, followed by those who took one test, while those who had simply re-read the passage remembered the least. This pattern of results is

remarkable because those with the greatest long-term retention of information had had the least exposure to the source materials. Those students had only read the materials once, and otherwise depended on their own recall in the free-response tests without the ability to check their answers for accuracy.

In medical education, one of the first studies to examine retrieval practice using an educationally relevant timeframe involved residents learning two topics: myasthenia gravis and status epilepticus.⁷ After an interactive didactic session, residents were randomized in a counterbalanced fashion to either take a short-answer test with feedback over one of the topics or study a review sheet over the other topic. Residents then took tests and studied the review sheets on two additional occasions each separated by 2 weeks. A final test on both topics was given 6 months later. As would be predicted by the direct testing effect, those who learned the topic through retrieval practice recalled an average of 13% more at 6 months than those who had repeatedly studied the topic. Importantly, 91% of the residents who responded to the end-of-study survey reported that they would be willing to take regular tests to improve their long-term retention.

The direct testing effect has been widely replicated in many settings with a wide range of populations, materials, and timeframes. Studies have demonstrated improved memory through retrieval practice with medical students, practicing physicians taking CME courses, middle school students, and undergraduate college students.^{8–11} Retrieval practice has not only shown improved memory of factual information but also of spatial relationships and procedural skills.^{8,12,13} Patients with multiple sclerosis, traumatic brain injury, and aphasia have all shown benefits in learning with retrieval practice.^{14–16} Retrieval practice has been shown to be superior in producing long-term retention compared with other study methods such as concept mapping and students writing explanations.^{17,18} A review of educational interventions based on cognitive psychology research found retrieval practice to be broadly supported in the education and psychology literature.¹⁹ Out of a list of evidence-based education recommendations by the United States Department of Education, two of the seven were based on retrieval practice.²⁰

How Does Retrieval Practice Work?

Many possible theories have been suggested for how retrieval practice might produce its mnemonic effects. One theory is the elaboration hypothesis which posits that the act of retrieval causes a person to construct more extensive relational networks between items.²¹ An alternative view is the episodic context hypothesis which asserts that retrieval causes a person to improve their recollection of contextual details as they search for the correct answer to a question.²² Despite these possibilities, no definitive evidence has emerged to directly answer the question of how retrieval practice works. Each proposed theory involves some sort of schema formation in which connections are made between elements to create and strengthen memory. It is likely that

each theory may work for particular types of materials or that a combination of theories best explains the effects of retrieval practice.

Retrieval practice depends on the basic processes of memory. The concept of memory consolidation (with the associated concept of reconsolidation) is likely a key factor in how retrieval practice works.^{23,24} *Consolidation* is the process by which a memory becomes enduring over the long term.^{24–26} To understand consolidation, it is critical to recognize that the processes of encoding, consolidating, and retrieving memories is not a static process of information storage but rather a constructive and reconstructive process in which memory is constantly changing.^{24–26} Memory processing involves encoding as a representation of the stimulus is created in the brain. For declarative (fact-based) memories, this process typically involves both the hippocampus and neocortex.^{24,25} The memory then undergoes consolidation over time as the representation becomes reorganized and distributed within the neocortex.^{23–26} This process is thought to occur through ongoing interactions between the hippocampus and the neocortex resulting in a memory that is less dependent on the hippocampus.^{23–26} Consolidation occurs as memories are reactivated in the cortex and integrated with preexisting circuits.^{23–26} Reactivation and replay often occur spontaneously while at rest and during sleep.^{23–25} Some evidence has emerged to suggest that retrieval practice leads to more rapid consolidation.²³ For instance, in one retrieval practice study investigating the role of sleep, retention for material that was restudied improved with sleep, approximating or even eliminating the advantage in retention gained through testing.²⁷ However, retention from retrieval practice did not improve with sleep, potentially indicating that consolidation had already occurred. This result may come from the external circuit reactivation that comes with retrieval practice.

While consolidation occurs over time, the speed at which a memory is consolidated seems to be dependent on the strength of the schema to which it is integrated.^{24,25} With greater schema strength comes more rapid consolidation. Schema integration could be another potential role for retrieval practice. Free recall tests have been shown to improve schema construction.²⁸ The act of retrieval may cause learners to construct schemas or may better integrate information into existing schemas.²⁹

As memories are retrieved, they can become malleable and are updated and reprocessed with novel information, thus strengthening the memory further.²⁴ This process of reactivating and updating memory is known as *reconsolidation*.²⁴ During the process of reconsolidation, memories can be influenced by various factors such as emotional cueing. As evidence that retrieval practice induces reconsolidation, the act of observing a negatively charged image, such as a dead animal or a person pointing a gun, after retrieving information led to improved retention of that information.³⁰ Seeing the image just prior to retention did not lead to the same improvement in retention (i.e., the effect did not arise simply because the image made the retrieval event more unique but rather influenced the reconsolidation processes that only occur after

retrieval).³¹ Viewing the negatively charged image after restudying did not lead to improved retention, suggesting that restudying information did not elicit reconsolidation.³⁰ Interestingly, viewing positive images did not improve retention.³¹

Functional imaging studies have shown variable results regarding the anatomical correlates of retrieval practice.^{32–36} While the patterns have been different between studies, structures typically thought to play a part in retrieval such as the hippocampus, prefrontal cortex, temporal lobe, and parietal lobe have been identified.^{26,33–36} An important finding in some functional imaging studies is that the areas of brain activation in retrieval practice are different than those in restudying.^{33,34} A limitation of the functional imaging research of retrieval practice is that most studies typically use simple materials and retrieval intervals on the order of days and weeks rather than months or years. Because consolidation processes change over time, it would be important for studies of retrieval practice using longer intervals to be conducted in the future.³⁷ More research is needed to better define and understand the anatomical systems involved in retrieval practice.

How is Retrieval Practice Best Implemented?

While the evidence supporting retrieval practice is robust, the take home message is not necessarily that more and more tests need to be added to curricula. Simply adding tests may not result in the desired mnemonic effect. Indeed, negative studies of retrieval practice have been published as well.^{38,39} Principles have emerged from the psychology and education literature that can guide implementation of retrieval practice to increase the likelihood that long-term retention is achieved. Four of these are considered below: test format, repetition, spacing, and feedback (– Fig. 1).

Test Format

As mentioned above, incorporation of information into a schema seems to be an important factor in the speed of memory consolidation.^{24,25} Educators should consider this principle as they design test formats. Tests that enhance learners' schema formation and schema reactivation may have the greatest benefit in long-term retention. Free recall tests have been shown to increase schema formation.²⁸ Tests that demand less schema construction or reactivation from the learner may not produce as great of an effect. In one study with first-year medical students, when standardized patient encounters using neurological diseases were compared with short answer tests which were both compared with restudying, the standardized patient encounters produced greater long-term retention than both short-answer tests or restudying when the final test consisted of a standardized patient encounter 6 months later.⁴⁰ When the final test was a short-answer test, both repeated retrieval through standardized patient encounters and short-answer tests performed equivalently and both were better than restudying. In this case, the standardized patient encounter could be seen as a free recall test. In the short-answer test, the very presence of questions creates a structure for the learner that does not

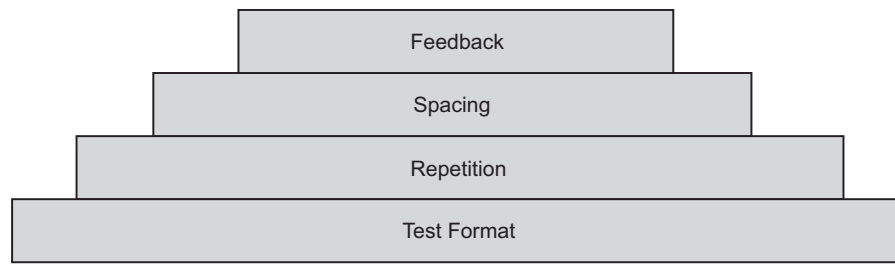


Fig. 1 Foundational blocks of retrieval practice implementation. Successful implementation of retrieval practice is not based solely on adding tests to educational experiences. Rather, educational implementation of retrieval practice is more likely to be successful when built on a foundation of test formats that encourage broad schema activation, providing adequate repetition that is spaced over weeks and months, and giving feedback to correct errors.

require them to create and reactivate as extensive of a schema for themselves. When the supports of that structure provided by the questions are removed the learner may not perform as well. A similar pattern of results was seen in a study of short texts, in which free recall tests produced better retention than fill-in-the-blank tests, which did better than true/false tests which were better than controls who had no further exposure to the texts.⁴¹ This pattern was the same regardless of the type of final test (free recall, fill-in-the-blank, or true/false). The above studies illustrate a complementary principle that greater retrieval effort in the practice tests produces greater long-term retention.^{42,43} Retrieval effort may simply be a marker for the extent of schema creation and reactivation.

Because multiple-choice questions (MCQs) are so common in educational practice, they deserve special mention. When MCQs are used for assessment purposes, they have been shown to have equivalent discriminating power as more open-ended questions,⁴⁴ and they make scoring much more convenient for educators. However, the use of MCQs in retrieval practice to improve long-term retention is more complicated. In some cases, MCQs have been shown to produce no more long-term retention than restudying when compared with short-answer tests.^{45,46} In other contexts, retrieval practice with MCQs has improved long-term memory.^{46,47} Whether or not MCQs are effective in retrieval practice likely relates to how they are designed. If the question consists of recognizing a simple fact then it likely does not produce much retrieval effort nor does it reactivate much of a schema and may not result in improved long-term retention. However, MCQs can be designed to require multiple steps of reasoning and more extensive schema activation.⁴⁶ Also, some reactivation and learning may occur for associated items among the lures in the MCQ.⁴⁸ Some evidence shows that these types of questions can produce equivalent retention to short-answer questions.⁴⁶ Even when MCQs replicate complex thought processes such as clinical reasoning, learners still engage in strategies of looking for clues in the question and formatting to help them find the answers.⁴⁹ MCQs can be used for retrieval practice but should be used with care. Other test formats may better accomplish educators' aims.

As educators design systems to incorporate retrieval practice, they should consider the breadth of potential retrieval

activities at their disposal. As mentioned above, standardized patients or other simulations offer powerful opportunities for retrieval. Repeated practice of procedural skills using simulation with mastery learning has shown extremely high levels of retention even after 12 months.⁵⁰ Real patient encounters are also opportunities for retrieval. For instance, when a neurology resident sees a patient with a new diagnosis of multiple sclerosis, she or he has to retrieve the diagnostic criteria, the risk factors, the expected exam findings, and the treatments. Various written test formats have been shown to lead to successful retrieval practice. One such approach is key features testing, in which learners answer clinical reasoning questions as they are presented with a narrative case.⁵¹ Essay questions are essentially free recall questions that can be used to encourage learners to engage in extensive retrieval. Using variation in questions may help in the creation of transferrable schemas. For example, when learners answered varied questions they used more of the information in a novel application test compared with if they were exposed to the same question repeated multiple times.⁵² As a guiding principle, educators should consider the extent of the schemas that they want learners to create and reactivate. They should provide retrieval opportunities that support schema formation for long-term retention. This approach likely means avoiding question and formats that require only isolated facts to be retrieved and rather using formats that require learners to generate an organizational structure around the information (→ Fig. 2).

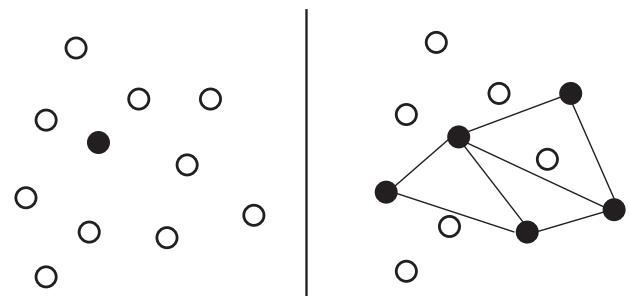


Fig. 2 Schema retrieval and activation. Each dot represents an element of a memory. Darkened dots represent items retrieved. Lines represent relationships between items. Retrieval opportunities are likely to be the most educationally effective when they require schema retrieval or reactivation as represented on the right as opposed to isolated fact recall as illustrated on the left.

Repetition

When examined over relatively long retention intervals, retrieval practice shows a dose effect in which more retrieval practice leads to better retention.⁵³ However, after several repetitions a point of diminishing returns is reached in which only small gains occur with additional retrieval practice.⁵³ For example, in one study using psychology terms, little additional benefit was seen after more than three successful retrievals.⁵³ When tests are closely spaced (discussed below), repetition does not clearly provide a benefit and one test may be as effective as multiple.^{43,53} If logistics force educators to choose between a single test versus no test, having at least one opportunity for retrieval practice may be more beneficial than no retrieval practice depending on the retrieval interval.

Repeated retrieval practice facilitates the memory reactivation and updating that is a critical part of consolidation and reconsolidation. When testing is combined with feedback (discussed below), learners are able to use repetition to make corrections and have opportunities to correctly retrieve information that they may have missed before. The process of updating memories also may explain the diminishing returns of too frequent repetition. As retrieval is no longer perceived as novel, the memory may not further consolidate.²⁴ Also, the diminishing returns may simply be a function of ceiling effects as memory capacity is not infinite. The number of repetitions that is needed for effective long-term retention likely depends on the quantity and complexity of the material to be learned. More complex and extensive materials may require more repetitions than simpler materials.

Spacing

Repetition cannot be discussed without also considering spacing. Spacing is a powerful, but often neglected principle of long-term learning. Much cognitive psychology research has shown that spacing improves learning.¹⁹ Spacing improves retention from restudying as well as retrieval. Spacing has also improved learning in surgical skills and medical knowledge.^{54,55} The effect of spacing depends on the retention interval.⁵⁶ Short spacing intervals lead to improved retention over a short retention interval. Long spacing intervals are needed to retain information for longer periods of time. For information to be retained for months or years, retrieval should be spaced on the order of weeks or months. In one study of medical students learning nephrology facts, when students took four tests back to back, retention was superior compared with restudying at 1 week.⁵⁷ However, at 6 months the level of retention for information that was tested was no different than information that had simply been restudied. Experiments which have shown improved retention over 6 months used retrieval intervals of 1 to 2 weeks.^{7,9,18,40} The improved short-term retention with close spacing intervals may explain the effectiveness of cramming in preparing for imminent tests—and why that information is quickly forgotten.

The effects of spacing are likely due to the processes of consolidation. Over the course of weeks, months, and years, consolidation leads to changes in memories as they are

reorganized in the neocortex.²⁵ Spacing may have its effect by reactivating and strengthening these more distributed and integrated circuits. As time passes and details are forgotten, it is possible that spacing may cause the retrieved stimulus to be perceived as novel and therefore lead to updating of the consolidated memory.

One challenge with spacing may occur if the retrieval interval is too long and so much forgetting may have occurred that extensive relearning must happen to achieve an acceptable level of performance. However, as discussed above, if spacing intervals are too close, then retrieval may not have the desired long-term effects. A potential strategy to find the appropriate balance is to use expanding retrieval intervals.⁵⁸ Short intervals can be used at first to maximize learning the details of the materials. The intervals are then spaced further out over time to optimized long-term retention.

Feedback

For retrieval practice to improve long-term retention of information, feedback is not necessary. Research has illustrated that the act of retrieval itself even without feedback produces a mnemonic benefit.⁶ However, studies have also shown that feedback can dramatically amplify the improvement of retention achieved with retrieval practice. In one study of general knowledge facts, participants recalled 41% of information on a final test that was previously tested without feedback compared with 24% of control items that were not tested.⁵⁹ Participants recalled 87% of information that was tested with feedback. For educational purposes, there is no benefit to withholding feedback. Generally for retrieval practice to be effective, retrieval must be successful.⁴³ However, difficult materials that require effortful retrieval may lead to increased unsuccessful retrieval attempts. Feedback can be useful to overcome the lack of benefit with unsuccessful retrieval.⁴³ Interestingly, when a learner gives an incorrect answer but is provided with corrective feedback, researchers have elicited evidence of reconsolidation.³¹ However, if a learner does not attempt to answer a question and corrective feedback is given, researchers have not found evidence of reconsolidation. These findings emphasize how at least attempted retrieval reactivates memory and allows it to be updated; therefore, guessing may be beneficial rather than simply skipping a question—especially when feedback is given.

Feedback is important for metacognitive monitoring. As learners use feedback to identify and correct errors, they make changes to the retrieval strategies that allow them to improve retention. In one set of experiments, students learned foreign language vocabulary through the “keyword” method in which learners select a word from their own language to help them remember the meaning of the foreign language word.⁶⁰ Not only were words learned through retrieval practice better retained, but learners also changed their keyword more frequently for words learned through retrieval practice with feedback than words learned through restudying. Retrieval practice with feedback provides a means by which learners can accurately assess their performance and modify their approach as needed.

Challenges and Opportunities in Using Retrieval Practice

The integration of retrieval practice into an actual curriculum presents many challenges (as well as opportunities!). One of the first challenges that educators face is deciding what to test. Everything within a curriculum cannot be tested—especially on a repeated basis. Given this fact, test formats that require the retrieval of sets of information either through schema production or multiple levels of processing become particularly important to maximize the extent to which the tests can cover material in the curriculum. Fortunately, evidence exists that information that is related to material that is retrieved but that is not in itself tested can show improved retention through testing as well.^{61–63} This phenomenon is known as *retrieval-induced facilitation*.⁶¹ Some educators may be concerned that courses will “teach to the test.” However, retrieval practice forces educators to consider which elements of the curriculum should be prioritized. If tests accurately represent these priorities and produce long-term retention, then the overall objective of the educational activity is met.

The idea of creating and grading repeated tests for large numbers of learners may discourage educators from implementing retrieval practice. However, some test formats such as free recall tests may reduce the work of test creation compared with multiple-choice tests which require significant effort to write well and to create viable alternative answers. Educators may be concerned about the effort in scoring essay and short-answer tests. However, when tests are used for learning and not for summative evaluation, the scoring can be performed by students. For free recall tests or other test formats, educators can provide model answers which students can use for feedback and scoring. Educators could score a sample of tests to ensure that students use the feedback resources appropriately and make expected progress.

Educators and academic leaders might wonder if the work required to create materials and systems for retrieval practice is warranted given that students seem to do well on final course examinations. Indeed, scores on course examinations often seem much higher than the percentages of retention measured in many retrieval practice research studies. As mentioned above, the actual long-term retention of most learners is probably far below their performance on course examinations because cramming leads to short-term retention.⁵⁷ Standard course examinations largely measure students’ cramming ability rather than durable learning. As such, the traditional education system creates an illusion of learning success without supporting and incentivizing strategies for long-term retention.

When educators begin to plan for long-term retention, retrieval practice provides an opportunity for longitudinal integration of curricula. Repeated, spaced retrieval practice allows earlier elements of the curriculum to be brought forward and retained. Creative test construction would provide opportunities to apply elements taught previously in the curriculum to situations and principles being taught

later in the curriculum. Through retrieval practice, educators will have a clear sense of the strengths and weaknesses of their learners over time.

Despite the strong evidence for the efficacy of retrieval practice, though, it should be acknowledged that the principles outlined in this article pertain to acquiring and retaining fact-based knowledge. This type of knowledge is only one type of learning.⁶⁴ Educational systems must also acknowledge and optimize learning that comes only through experience and the social interactions of authentic work. Retrieval practice is only one tool that can be used in a complex educational system.

Conclusion

As educators seek for evidence-based interventions to create durable learning, retrieval practice has robust support in the cognitive psychology and education literature. Retrieval practice likely generates its effects through the memory consolidation and reconsolidation processes in the brain. As educators design retrieval practice systems, they should consider test formats that activate broad schemas of information and require effortful retrieval. Educators should provide adequate opportunities for repetition that is spaced over weeks and months to lead to retention that can last for months and years. Finally, tests should include feedback to ensure that errors are corrected and to enable learners to identify the most effective retrieval strategies. By applying these principles, educators can use retrieval practice to plan for and facilitate long-term retention of knowledge.

Glossary of Key Terms

Direct testing effect—The phenomenon in which information that is retrieved as part of taking a test is better retained than information that is simply studied.

Retrieval practice—The process of retrieving information to improve retention.

Encoding—The process by which a stimulus is transformed into a memory in the brain.

Consolidation—The process by which a memory is reorganized and integrated into the neocortical circuits so as to become less dependent on the hippocampus.

Reconsolidation—The reactivation of consolidated memories for updating and strengthening.

Schema—A network of related information.

Retrieval-induced facilitation—The phenomenon of improved retention of information that is related to material that is retrieved but that is not itself tested.

Acknowledgment

The author would like to thank Andrew Butler for his discussions and insights on the topic, especially his thoughts on consolidation.

References

- 1 Roediger HL III, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. *Perspect Psychol Sci* 2006;1(03):181–210

- 2 Roediger HL III, Butler AC. The critical role of retrieval practice in long-term retention. *Trends Cogn Sci* 2011;15(01):20–27
- 3 Larsen DP, Butler AC. Test-enhanced learning. In: Walsh K, ed. *Oxford Textbook of Medical Education*. Oxford: Oxford University Press; 2013:443–452
- 4 Gates AI. Recitation as a factor in memorizing. *Arch Psychol* 1917; 6:1–104
- 5 Jones HE. Experimental studies of college teaching: the effect of examination on the permanence of learning. *Arch Psychol* 1923; (10):1–70
- 6 Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 2006;17(03): 249–255
- 7 Larsen DP, Butler AC, Roediger HL III. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 2009;43(12):1174–1181
- 8 Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ* 2009;43(01):21–27
- 9 Larsen DP, Butler AC, Aung WY, Corboy JR, Friedman DI, Sperling MR. The effects of test-enhanced learning on long-term retention in AAN annual meeting courses. *Neurology* 2015;84(07):748–754
- 10 McDaniel MA, Agarwal PK, Huelser BJ, McDermott KB, Roediger HL III. Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. *J Educ Psychol* 2011;103:399–414
- 11 Karpicke JD, Roediger HL III. The critical importance of retrieval for learning. *Science* 2008;319(5865):966–968
- 12 Carpenter SK, Kelly JW. Tests enhance retention and transfer of spatial learning. *Psychon Bull Rev* 2012;19(03):443–448
- 13 Wayne DB, Siddall VJ, Butter J, et al. A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Acad Med* 2006;81(10, Suppl):S9–S12
- 14 Sumowski JF, Chiaravalloti N, DeLuca J. Retrieval practice improves memory in multiple sclerosis: clinical application of the testing effect. *Neuropsychology* 2010;24(02):267–272
- 15 Sumowski JF, Wood HG, Chiaravalloti N, Wylie GR, Lengenfelder J, DeLuca J. Retrieval practice: a simple strategy for improving memory after traumatic brain injury. *J Int Neuropsychol Soc* 2010;16(06):1147–1150
- 16 Middleton EL, Schwartz MF, Rawson KA, Garvey K. Test-enhanced learning versus errorless learning in aphasia rehabilitation: testing competing psychological principles. *J Exp Psychol Learn Mem Cogn* 2015;41(04):1253–1261
- 17 Karpicke JD, Blunt JR. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 2011; 331(6018):772–775
- 18 Larsen DP, Butler AC, Roediger HL III. Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ* 2013;47(07):674–682
- 19 Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychol Sci Public Interest* 2013;14(01):4–58
- 20 Pashler H, Bain P, Bottge B, et al. *Organizing Instruction and Study to Improve Student Learning: A Practice Guide* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education; 2007
- 21 Carpenter SK. Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *J Exp Psychol Learn Mem Cogn* 2009;35(06):1563–1569
- 22 Whiffen JW, Karpicke JD. The role of episodic context in retrieval practice effects. *J Exp Psychol Learn Mem Cogn* 2017;43(07): 1036–1046
- 23 Antony JW, Ferreira CS, Norman KA, Wimber M. Retrieval as a fast route to memory consolidation. *Trends Cogn Sci* 2017;21(08): 573–576
- 24 Dudai Y. The restless engram: consolidations never end. *Annu Rev Neurosci* 2012;35:227–247
- 25 Dudai Y, Karni A, Born J. The consolidation and transformation of memory. *Neuron* 2015;88(01):20–32
- 26 Ben-Yakov A, Dudai Y, Mayford MR. Memory Retrieval in mice and men. *Cold Spring Harb Perspect Biol* 2015;7(12):a021790
- 27 Bäuml KH, Holterman C, Abel M. Sleep can reduce the testing effect: it enhances recall of restudied items but can leave recall of retrieved items unaffected. *J Exp Psychol Learn Mem Cogn* 2014; 40(06):1568–1581
- 28 Zaromb FM, Roediger HL III. The testing effect in free recall is associated with enhanced organizational processes. *Mem Cognit* 2010;38(08):995–1008
- 29 Hinze SR, Wiley J, Pellegrino JW. The importance of constructive comprehension processes in learning from tests. *J Mem Lang* 2013;69:151–164
- 30 Finn B, Roediger HL III. Enhancing retention through reconsolidation: negative emotional arousal following retrieval enhances later recall. *Psychol Sci* 2011;22(06):781–786
- 31 Finn B, Roediger HL III, Rosenzweig E. Reconsolidation from negative emotional pictures: is successful retrieval required? *Mem Cognit* 2012;40(07):1031–1045
- 32 Eriksson J, Kalpouzos G, Nyberg L. Rewiring the brain with repeated retrieval: a parametric fMRI study of the testing effect. *Neurosci Lett* 2011;505(01):36–40
- 33 Wing EA, Marsh EJ, Cabeza R. Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia* 2013;51(12):2360–2370
- 34 van den Broek GSE, Takashima A, Segers E, Fernández G, Verhoeven L. Neural correlates of testing effects in vocabulary learning. *Neuroimage* 2013;78:94–102
- 35 Liu XL, Reder LM. fMRI exploration of pedagogical benefits of repeated testing: when more is not always better. *Brain Behav* 2016;6(07):e00476
- 36 Hashimoto T, Usui N, Taira M, Kojima S. Neural enhancement and attenuation induced by repetitive recall. *Neurobiol Learn Mem* 2011;96(02):143–149
- 37 Furman O, Mendelsohn A, Dudai Y. The episodic engram transformed: time reduces retrieval-related brain activity but correlates it with memory accuracy. *Learn Mem* 2012;19(12):575–587
- 38 McConnell MM, Azzam K, Xenodemetropoulos T, Panju A. Effectiveness of test-enhanced learning in continuing health sciences education: a randomized controlled trial. *J Contin Educ Health Prof* 2015;35(02):119–122
- 39 Mathes EF, Frieden IJ, Cho CS, Boscardin CK. Randomized controlled trial of spaced education for pediatric residency education. *J Grad Med Educ* 2014;6(02):270–274
- 40 Larsen DP, Butler AC, Lawson AL, Roediger HL III. The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Adv Health Sci Educ Theory Pract* 2013;18(03):409–425
- 41 Glover JA. The “testing” phenomenon: not gone but nearly forgotten. *J Educ Psychol* 1989;81:392–399
- 42 Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? *J Mem Lang* 2009;60:437–447
- 43 Rowland CA. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull* 2014;140 (06):1432–1463
- 44 Norman GR, Smith EKM, Powles ACP, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. *Med Educ* 1987;21(04):297–304
- 45 Butler AC, Roediger HL III. Testing improves long-term retention in a simulated classroom setting. *Eur J Cogn Psychol* 2007; 19:514–527
- 46 McConnell MM, St-Onge C, Young ME. The benefits of testing for learning on later performance. *Adv Health Sci Educ Theory Pract* 2015;20(02):305–320
- 47 McDermott KB, Agarwal PK, D'Antonio L, Roediger HL III, McDaniel MA. Both multiple-choice and short-answer quizzes enhance

- later exam performance in middle and high school classes. *J Exp Psychol Appl* 2014;20(01):3–21
- 48 Little JL, Bjork EL, Bjork RA, Angello G. Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychol Sci* 2012; 23(11):1337–1344
 - 49 Surry LT, Torre D, Durning SJ. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Med Educ* 2017;51(10):1075–1085
 - 50 Barsuk JH, Cohen ER, McGaghie WC, Wayne DB. Long-term retention of central venous catheter insertion skills after simulation-based mastery learning. *Acad Med* 2010;85(10, Suppl): S9–S12
 - 51 Raupach T, Andresen JC, Meyer K, et al. Test-enhanced learning of clinical reasoning: a crossover randomised trial. *Med Educ* 2016; 50(07):711–720
 - 52 Butler AC, Black-Maier AC, Raley ND, Marsh EJ. Retrieving and applying knowledge to different examples promotes transfer of learning. *J Exp Psychol Appl* 2017;23(04):433–446
 - 53 Rawson KA, Dunlosky J. Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *J Exp Psychol Gen* 2011;140(03):283–302
 - 54 Moulton CAE, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial *Ann Surg* 2006;244(03): 400–409
 - 55 Kerfoot BP, Baker HE, Koch MO, Connelly D, Joseph DB, Ritchey ML. Randomized, controlled trial of spaced education to urology residents in the United States and Canada. *J Urol* 2007;177(04): 1481–1487
 - 56 Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. Spacing effects in learning: a temporal ridge of optimal retention. *Psychol Sci* 2008;19(11):1095–1102
 - 57 Schmidmaier R, Ebersbach R, Schiller M, Hege I, Holzer M, Fischer MR. Using electronic flashcards to promote learning in medical students: retesting versus restudying. *Med Educ* 2011;45(11): 1101–1110
 - 58 Dobson JL. Effect of uniform versus expanding retrieval practice on the recall of physiology information. *Adv Physiol Educ* 2012;36 (01):6–12
 - 59 Butler AC, Karpicke JD, Roediger HL. Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *J Exp Psychol Learn Mem Cogn* 2008;34(04):918–928
 - 60 Pyc MA, Rawson KA. Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *J Exp Psychol Learn Mem Cogn* 2012;38(03):737–746
 - 61 Chan JCK, McDermott KB, Roediger HL III. Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. *J Exp Psychol Gen* 2006;135(04): 553–571
 - 62 Chan JCK. Long-term effects of testing on the recall of nontested materials. *Memory* 2010;18(01):49–57
 - 63 Rowland CA, DeLosh EL. Benefits of testing for nontested information: retrieval-induced facilitation of episodically bound material. *Psychon Bull Rev* 2014;21(06):1516–1523
 - 64 Hager P. Learning and metaphors. *Med Teach* 2008;30(07):679–686