14

Strategies to Improve Learning and Retention During Training

HENRY L. ROEDIGER, III, JOHN F. NESTOJKO,
AND NICOLE S. SMITH

INTRODUCTION

Training is one of the most critical operations of every organization. However, training is more critical in military service than in practically any other endeavor. Proper training is critical to national security and can be a life and death matter for those being trained. If some members of a unit are poorly trained, all members of the unit may be endangered.

In the US military services, Navy SEALs have the longest and most arduous process of training. The basic training lasts for more than a year, and around 75% of those beginning training drop out along the way. To be allowed to begin training, a person must pass a challenging set of physical and mental tests, so only the hardy and prepared even begin SEAL training. The first step to becoming a SEAL is to complete 8 weeks at the Naval Warfare Preparatory School at Great Lakes, Illinois. Those who make it through this course then begin real SEAL training. The next phase involves 24 weeks of Basic Underwater Demolition/SEAL (or BUD/S) training: 3 weeks of indoctrination training, then 7 weeks of physical conditioning, then 7 weeks of combat diving, then 7 weeks of land warfare. Those making it through this first phase of SEAL training (and many do not), then go on to Parachute Jump School (3 weeks). After that, they have 26 weeks of SEAL Qualification Training. Graduating from this course earns the prized Navy SEAL Trident. Only after this step is a SEAL assigned to a unit (e.g., SEAL Team 4). However, training has not ended. Navy SEALS continue higher level training throughout their career. Anyone interested in how Navy SEALS are trained should read Dick Couch's two fascinating books (2004, 2009) on the subject; a quicker overview can be found in the Wikipedia entry on Navy SEAL selection and training at https://en.wikipedia.org/wiki/United_States_Navy_SEAL_selection_and_training.

Because of the rigors of SEAL training, the public, as well as every branch of military service, regard the training of SEALs as the pinnacle of military training in the entire world. Their motto is "The only easy day was yesterday." Their training is the most difficult, as well as the longest, of any military organization.

Given this backdrop, the first author of this chapter was stunned to receive an email message from Carl Czech, one of the men responsible for selecting and training Navy SEALs, directed to the authors of *Make It Stick: The Science of Successful Learning* (Brown, Roediger, & McDaniel, 2014). With permission, we provide his note (from July, 2016) here:

Gentlemen,
I'm an instructional systems specialist and advisor at the Naval Special Warfare Center in Coronado, California. We and our subordinate commands are responsible for the selection and training of Navy SEALs. Our programs encompass everything from basic combat skills to advanced special operations.

Since I joined the SEAL community a decade ago, my professional colleagues and I have endeavored to make research-based instructional design a central tenet of Naval Special Warfare training. I write to tell you how much your work has helped us progress along that path. Shortly after "Make it Stick" was published, my Commanding Officer at the time came to me and said, "You keep telling me about research on learning and instruction. Is there something you could give me to read? Something that summarizes it in a straightforward way?" Luckily, I had a copy of your book on my desk. I never saw it again. He read it, talked about it, and then passed it on to the commander who succeeded him a few months later.

Subsequently, we obtained and distributed dozens of more copies and have made it a touchstone for our instructors and support staff. In a community that depends a great deal on apprenticeship and craft knowledge, we're beginning to develop a new appreciation for serious learning research. By communicating solid cognitive psychology principles in elegant and meaningful ways, we've begun to improve our instructional designs and delivery. This is no mean feat, since the vast majority of our instructors are assigned from the operating forces for only a couple of years at a time.

Last month, we convened an informal, one-day meeting of instructional designers, instructors and staff leaders. As we brainstormed improvement ideas, "Make it Stick" provided sturdy scaffolding for lively and focused discussion. Our plan is to continue the process with follow-on meetings that will extend to the larger Naval Special Warfare organization, including the training cadres who prepare SEAL Teams for deployment.

I'd personally appreciate any advice you may have for us as we move forward. In turn, we'd be happy to share our experience in how we're trying to gain real advantage for our learners as they tackle some of the toughest work in the world. Your work has been a recent part of that. Thanks again, from the guys in the field.

Best Regards,
Carl
Dr. Carl Czech
Senior Instructional Systems Specialist/Advisor
Naval Special Warfare Center
Coronado, CA 92155

Really? The commanders and trainers of the Navy SEALs are reading our book explaining principles of cognitive psychology and finding that they are useful in changing the procedures by which Navy SEALs are trained? The answer was *yes*. What ensued were numerous email exchanges and several telephone calls in which we learned what advice in our book was most useful in changing SEAL training. Eventually the first author of this chapter and Mark McDaniel, the other cognitive psychologist author of *Make It Stick*, flew to San Diego and spent 2 days on Coronado Island meeting with Navy SEAL trainers. We made a presentation to perhaps 60 trainers, many of whom were themselves SEALs, as well as to Commander Jay Hennessey. The visit was enlightening on both sides.

In reflecting on the issues that were facing Navy SEAL trainers, we see that they are ones that are common to practically all forms of training, not just military training. The principles occur in sports training, in musical training, in business training, and in learning in educational settings.

In this chapter, we provide three central strategies to improve training. These were all ones that the SEAL trainers began implementing, where possible, after reading our book, and they are ones that are not commonly incorporated into training in the military, in sports, or in education. Before we get to the new strategies, though, we consider typical means of training and discuss reasons why trainers see it as more effective than it is. We will use a sports analogy—learning to hit a baseball—as an example.

A TYPICAL TRAINING REGIME

Training of complex skills and procedures of any sort occurs over an extended period. A college baseball player must learn to hit the baseball, field his position, run the bases, and learn the rules of the game. Obviously, anyone who can make a college team has already played baseball for many years and achieved a certain level of success. How to improve?

The commonly used technique is repetition, practicing the same skill over and over until a greater level of proficiency is achieved. If a batter has trouble hitting a curve ball, he can expect the coach to throw him perhaps 30 curve balls in a row in practice. Sure enough, he gets better at hitting the curve. Similarly, a third baseman who has trouble fielding bunts will get a long series of bunts to field. These techniques are called *massed training*, and they are routinely used in training of every sort. The reason is that this technique supports rapid learning; if someone practices the same skill over and over, he or she is thought to build up "muscle memory" so that the skill will become more automatic. (Muscle memory is not a term cognitive psychologists endorse, but it is widely used in training.) And massed repetition does build up fast learning, but it leads to a major problem: skills learned this way decay rapidly; forgetting is as rapid as learning in some cases.

What should be done instead? Training on a particular skill should be spaced out in time—*spaced practice*—not bunched up all at once. After all, pitchers do not telegraph their pitches; there will never be a case where a batter knows that every pitch in his at bat will be a curve ball. The third baseman never knows when the bunt is coming. Besides spacing, the other principle that is useful in training is interleaving. A baseball hitter never knows what is coming next (fastball? change up? slider?). A batting practice pitcher should interleave the types of pitches that a batter receives, just as will happen in a game. The mantra in sports is often, "Practice like you play, and you will play like you practice." However, this bit of (good) advice would indicate that many techniques used in practice—doing the same thing over and over—should be discontinued because that will never happen in a game.

To formalize the preceding points using the language of cognitive psychology, we can say that trainers should emphasize *transfer appropriate processing* (e.g., Roediger, Gallo, & Geraci, 2002)—that is, the type of process during learning or training should match the way the skill will be used after training. This term captures the "practice like you play" phrase. The trainer should always keep in mind how the skill to be learned will be deployed in the field and have the trainee practice that way, especially in later stages of training. If the processes used in training mimic those that will be needed after training, in a game or in combat, then that is the way they should be practiced. Once this is pointed out, it is easy to nod in agreement. However, most training regimes violate this principle because they emphasize massed practice.

As we shall see, evidence from much research supports the principles of spaced and interleaved training as producing good long-term retention. Why don't trainers use these techniques as a matter of course? The answer is that these strategies of training slow learning and make it feel (for both the trainer and the learner) that not much progress is being made. Massed practice is much more satisfying because the learner gets better faster, but these gains are often short-lived and illusory, fading rapidly and failing to support good long-term performance. When learners experience rapid acquisition that does not stick in the long-term, they suffer from illusions of competence—they believe their training was effective and that their performance will remain high in the long run, when in fact it only appeared effective during training (Bjork, 1994).

Why do learners—and trainers—believe in and continue to use training methods that fail to produce lasting results? A helpful distinction is between *performance* during acquisition and *learning* in the long run. Army recruits must learn the Soldier's Creed to perfection and are expected to recite it at critical points, such as when they graduate from basic training (a quick search online uncovers discussion forums full of new soldiers agonizing over how to learn the Soldier's Creed to avoid embarrassment during basic training). Now imagine a soldier who simply studies the words over and over, hoping they will sink in. After enough exposure, he finally finds that he can get it right. Satisfied with his apparent learning, the soldier puts it aside to focus on other training. A week later, his drill sergeant calls on him suddenly to recite the Soldier's Creed and the young recruit finds that the words simply won't come. What went wrong? Why can he no longer remember what he knew just a week ago? In this situation, the soldier mistook performance during acquisition (correctly reciting the Soldier's Creed once after crammed study) as an indicator of durable learning. The soldier neglected to consider a fundamental property of memory: forgetting over time. Learners often fail to consider how strongly forgetting will reduce performance in the future, assuming that if they know something at one point it will continue into the future. This is why performance during training can often be a poor indicator of long-term mastery (Soderstrom & Bjork, 2015). Massed practice, as noted earlier, gives the illusion of mastery, but other techniques are needed to make learning stick. It is exactly this misplaced trust in performance measured during acquisition—as opposed to evaluations that take place long after training—that explains why ineffective training strategies remain popular.

Let us introduce another term from cognitive psychology that helps us to understand this situation: *Desirable difficulties in learning* (Bjork, 1994). This idea captures the fact that in several situations, techniques that produce good short-term learning lead to poor long-term retention. Massed practice is an example. Rather, other strategies for training that are more difficult, feel bad to the learner (and trainer), and actually slow learning actually lead to better performance in the long term. Spaced practice, interleaved practice, and retrieval practice (e.g., via tests) are three of these desirable difficulties. All three feel somewhat unnatural or difficult when used during learning, but they lead to good retention when measured after a delay. We will present evidence supporting these claims in the next sections of the chapter.

Training also occurs in the classroom, for soldiers being trained, for athletes, and, of course, for traditional students. Baseball players go over rules for many game situations that happen infrequently (e.g., the infield fly rule). Or infielders must learn how to position themselves for relay throws from the outfielders in different situations (e.g., men on first and third and the ball hit to the center fielder either in the air or after a bounce). Similarly, the outfielders must learn where to throw to the cutoff man depending on different situations. Pitchers and catchers must learn to back up the plays depending on all these factors. Often these situations are presented in PowerPoint presentations after a practice, with the weary players expected to pick up the nuances.

"Death by PowerPoint" is an all too common experience for people in training. Often the slide shows are given out so that people can go through them repeatedly. Yet evidence shows that this form of study is ill-conceived and produces illusions of learning. That is, the trainees may know the information briefly, but they lose it quickly. Research shows that a much better technique for making the information stick for the long term is to practice retrieving it rather than simply being exposed to it repeatedly. That is, trainers should quiz trainees in the classroom and then put them in the practical situation where they need to perform and ask them to provide the answer; the trainee needs to practice retrieving or using the information to retain it for the long haul. Getting information out of memory is as critical to learning as is getting it into the memory system in the first place.

This section provides a quick overview of some of the topics in our chapter. The three principles we cover—spaced practice, interleaved practice, and retrieval practice—all produce durable and flexible learning. We now provide evidence to back up our claims. Although these principles apply across training situations—including training of first responders outside of the military—police, firefighters, and emergency medical technicians—we embed our principles in the context of problems that trainers confront in developing military personnel.

TESTING AS A LEARNING TOOL: THE POWER OF RETRIEVAL PRACTICE

Tests are part and parcel of any educational system. Students stay up all night studying to earn a passing grade, instructors and universities base course grades largely on test results, and standardized tests dictate the direction of people's lives. Yet tests are typically viewed as a *measure* of learning or ability, not a method to enhance learning. This view is not altogether wrong; of course, tests are useful tools to assess what people do and do not know. However, tests are also a uniquely potent device to promote flexible use of knowledge and to entrench knowledge deeply.

Tests represent a good example of transfer-appropriate processing. In other words, tests create a good match between conditions of training and the conditions that may arise during future use of the trained knowledge and skills. Take, for example, students who read and re-read an assigned chapter prior to an exam. When they read it a second time, the material feels familiar and the reading fluent, and that familiarity and fluency may lead the student to a feeling of comprehension, as if they know the content; this is what we referred to earlier as an illusion of mastery. But assessment tests rarely provide students with chapters and ask, "Can you comprehend what this chapter says?" Instead, tests ask students to recall information, or to apply knowledge to solve problems. In other words, re-reading presents a poor match to the later method of assessment. Practice that requires retrieval—in other words, tests—is much better matched to conditions that arise during future use, including assessments (tests). It is for this reason that tests as learning devices can better be conceived of as *retrieval practice* because practicing retrieval is the appropriate method to prepare for the demands that arise during assessment tests—namely, retrieving knowledge and facts from memory. And in practical situations where learned information is to be used, it must be readily retrieved.

The beneficial effects of testing on retention was empirically demonstrated early in the twentieth century (Gates, 1917), but it was largely ignored as a learning tool until relatively recently. In one study, Roediger and Karpicke (2006) compared testing to repeated studying as a learning technique. They gave college students short passages to read about the sun and sea otters. In one experiment, students followed their initial study of the passage either with three additional study trials (SSSS, when including the initial study session) or they took three consecutive tests after initial study (STTT). The tests were difficult: in a form of testing called *free recall*, participants were told to recall everything they could from the passage they had read. To assess longer term learning, a final free recall test was given to both groups to assess memory for the material. Half of the students received this test 5 minutes after the last trial in the study schedule, whereas the other half of the students received their final test a week later. When tested soon after learning, students who studied the passage four times (SSSS) recalled about 12% more content than did the students who studied the passage once followed by three tests during learning (STTT), an advantage of repeated studying (see Figure 14.1). However, when this assessment test was given at a delay of 1 week, the pattern completely reversed: repeated testing (STTT) produced 21% greater recall than the repeated studying condition (SSSS).

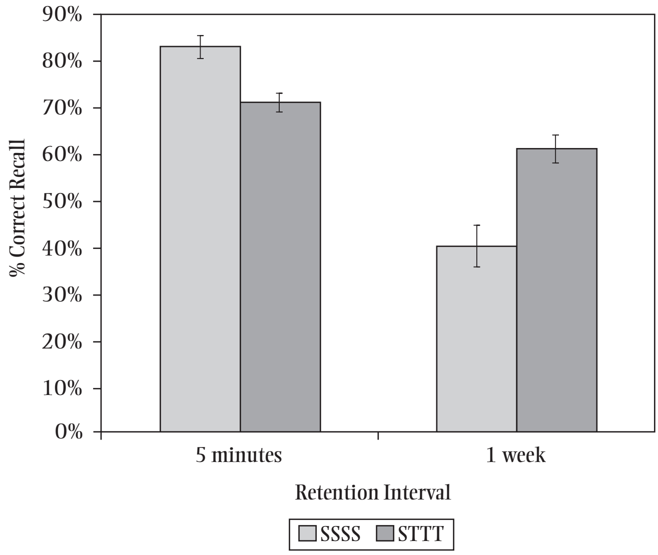

Figure 14.1 The results of Roediger and Karpicke (2006, Exp. 2). When tested 5 minutes after learning, students recalled more content from a text passage when they repeatedly studied (S) versus repeated testing (T); however, delayed recall was superior following repeated testing, demonstrating that tests enhance long-term retention. Error bars represent standard errors of the means and were recreated from the original figure. Adapted from H. L. Roediger & J. D. Karpicke, Test-enhanced learning. *Psychological Science, 17,* 249–255, 2006.

This effect showing that cramming (repeated studying) produces rapid forgetting can be observed by comparing the 5-minute to the 1-week delayed tests in the figure: participants forgot nearly half of what they could recall when they initially studied four times (SSSS), but they only forgot about 14% of the content over a week when they studied once and took three tests during learning (STTT). The primary take-home from this experiment is that tests enhance long-term retention of information by slowing forgetting.

The enhanced retention following repeated retrieval practice is especially impressive given that the students spent considerably more time with the passage in the repeated study schedule than in the repeated testing schedule (i.e., studying four times compared to studying only once), so they had more opportunities in that condition to learn with the material right in front of them. Furthermore, students were unaware of the benefits of testing in this experiment. When asked how much of the passage they would remember in 1 week just after the learning phase of the experiment, the students who had studied four times (SSSS) predicted better delayed test performance than students who had studied once and taken three tests (STTT). Their predictions were exactly the opposite of the recall pattern on the tests given a week later. What went wrong? Perhaps learners are not skilled at predicting long-term forgetting, but instead make their predictions based on what they know when they make them. After all, their predictions were accurate regarding the tests given soon after learning. In other words, they mistook current performance to be an indicator of long-term learning, and it is this mistake that creates the illusion of mastery that is a hallmark of poor training techniques.

There are hundreds of experiments demonstrating advantages of retrieval practice with verbal materials like word lists, textbook chapters, vocabulary learning, and foreign-language acquisition (Adesope, Trevisan, & Sundararajan, 2017). We will limit our coverage, however, mostly to demonstrations of retrieval-enhanced learning in contexts we see as more fitting to military training.

All soldiers are trained on how to navigate terrain by using maps. As the following example reveals, tests during training enhance learning of maps. Carpenter and Pashler (2007) had students learn maps containing a dozen features such as roads, rivers, and buildings. Students attempted to learn the maps via a computer display either only through studying or through a combination of studying and testing. In the study-only condition, they simply viewed the map for 2 minutes. In the test-study condition, which also lasted a total of 2 minutes, they initially viewed the map with all 12 features in place for 20 seconds. After viewing the intact map for 20 seconds, they saw a version of the map with one feature missing, were asked to visualize the missing feature, and then were shown the missing feature along with the rest. They went through 12 trials like this, with each of the 12 features missing during one trial. Following the study-only and the test-study conditions of learning, students were asked to draw the maps from memory after a 30-minute delay. Students' drawings were more complete and more accurate following the test-study procedure than the study-only procedure, demonstrating that tests can enhance visual-spatial map learning.

Retrieval Practice Promotes Flexible Application of Learning

If the only benefit of retrieval practice was to enhance the durability of learning, we would still be advising its widespread use in education and training. We are especially strong advocates of the use of testing in training, however, because there are additional benefits beyond enhanced retention. One additional benefit is that testing often leads to greater flexibility of knowledge use.

Retrieval practice improves students' abilities to answer inference questions on assessment tests. Inference questions are test questions that require a student to use reasoning from the facts they learned as opposed to directly recalling the facts. For example, McDaniel, Howard, and Einstein (2009) had students initially study complex materials about the mechanical workings of brakes and pumps then gave them problem-solving questions that required the students to make inferences from the knowledge they gained during learning; they also received recall questions directly assessing individual fact retention. An example inference question was "What could be done to make brakes more effective, that is, to reduce the distance needed to stop?" The answer to this question was never directly presented in the text participants read; rather, they were required to piece together the answer from the facts they had learned. Students answered 14% more of these problem-solving questions correctly when they were tested during learning compared to when they only studied the material without tests. The benefit of retrieval practice was even larger for the recall of facts, demonstrating the typical direct benefit of testing.

Mechanical engineers in the military are responsible for designing and building complex machines like tanks and planes. Retrieval practice during training will enhance the retention and flexibility of their knowledge.

In another demonstration of flexible learning enhanced by testing—namely, transfer from one knowledge domain to another—Butler (2010) had students study text passages about one topic either by repeated reading or repeated testing, and then assessed their knowledge by giving them questions about a seemingly unrelated topic. For example, when an initial passage about bats included the fact that bat wings are more flexible than birds' wings, a transfer question was, "The US Military is looking at bat wings for inspiration in developing a new type of aircraft. How would this type of aircraft differ from traditional aircrafts like fighter jets?" (The answer: "Traditional aircrafts are modeled after bird wings, which are rigid and good for providing lift. Bat wings are more flexible, and thus an aircraft modeled on bat wings would have greater maneuverability"; Butler, 2010; p. 1127). Tests during learning were only about bats; the transfer questions were given on an assessment test 1 week after learning. Compared to the restudy condition, students who were repeatedly tested on their bat knowledge during learning answered the transfer questions correctly far more often than those who had re-read the material (68% vs. 44%, a whopping 24% difference).

Following up on the map-learning study just described (Carpenter & Pashler, 2007), Rohrer, Taylor, and Sholar (2012) found that tests enhance map learning involving novel tests of spatial knowledge. Briefly, they gave 4th and 5th grade children maps with 10 locations labeled. In a study-only condition, children labeled a paper map while viewing a screen with the labels in place (i.e., they simply copied the answers from the screen, which requires no retrieval from memory). In a test-study condition, they were given the name of a location and asked to identify it on a blank map, and then they were shown the correct answer on screen. In other words, they tried to retrieve the correct name and location before being shown the correct answer as feedback. One day later, they were then given a route to follow through the map, as if they were driving through the regions the map depicted, and asked to identify which locations they would "drive" through on their route. The test-study procedure during learning nearly doubled accuracy on this test relative to the study-only procedure.

Soldiers, like students, acquire knowledge in the classroom, but then they are required to apply that knowledge to novel problems that often occur in stressful, challenging situations. This type of flexible use of knowledge is critical to their success and survival. Retrieval practice during knowledge acquisition is a potent tool for enhancing the flexibility of knowledge use.

Additional Benefits of Testing

So far, we have outlined two benefits of retrieval practice: durability and flexibility of learning. Another benefit recently discovered is that retrieval practice guards against deleterious effects of stressful situations on memory (Smith, Floerke, & Thomas, 2016). Stress reduces memory retrieval (Gagnon & Wagner, 2016), a fact that can have disastrous consequences for soldiers. In the study by Smith and colleagues, participants initially learned stimuli either via repeated studying or repeated retrieval practice. One day later, their memory was tested either without stress induction or after a stress induction episode that required participants to give speeches and solve math problems in front of judges. The material learned under repeated studying showed the typical detriments of stress: recall was worse after stress induction than without stress induction. What happened to the material that was learned via repeated retrieval? In this case, stress did not decrease recall for the previously tested material, and previously tested material was recalled better overall than previously restudied material, the typical benefit of testing. Although the stress of giving a speech and solving math problems is not equivalent to the stress of jumping out of an airplane into combat, this demonstration provides a promising avenue for a method of reducing stress-induced retrieval failure.

Retrieval practice has a variety of additional benefits over nonretrieval methods of learning. For example, tests can identify gaps in a learners' knowledge, which can be useful for both trainees and trainers moving forward. Tests can also improve learners' own judgments of their learning, thereby reducing the illusions of competence outlined earlier. Retrieval practice promotes organization of knowledge, can facilitate retrieval of material that was not on the test, and can improve acquisition of subsequent learning. To discuss each of these benefits in turn would require a new chapter of its own, so we refer you to Roediger, Putnam, and Smith (2011), who discuss these and other advantages of testing.

Two Tips for Implementing Retrieval Practice: Feedback and Repeated Testing

There are numerous considerations when thinking of how best to use tests during training. What type of tests should one give? There are multiple-choice tests, essay tests, true/false tests, and so on. When should tests be given? How often? This list could be expanded. Here we suggest two best practices that we know make retrieval practice particularly effective (see Putnam, Nestojko, & Roediger, 2016, for a discussion of test format and other related issues).

Feedback

What happens when learners fail to recall the correct answer on a test question? At best, they fail to learn that particular fact or skill. At worst, if they produced an incorrect answer on the test, they could retain that wrong answer as if it is the correct one (Roediger & Marsh, 2005). An easy solution to this problem is to provide feedback after tests. According to three recent meta-analyses of the research on testing, feedback generally boosts the benefits of retrieval-based learning (Adesope, Trevisan, & Sundararajan, 2017; Rowland, 2014; Schwieren, Barenberg, & Dutke, 2017). The type and timing of feedback are important, though.

Feedback following tests can take a few different forms. Two popular types of feedback are verification feedback and answer feedback. Verification feedback consists of telling students whether they were right or wrong on a specific test item without providing the correct answer. There is evidence that this type of feedback does not improve performance relative to tests without feedback (Fazio, Huelser, Johnson, & Marsh, 2010; Pashler, Cepeda, Wixted, & Rohrer, 2005). In contrast, answer feedback consists of providing the correct answer after test trials, and this type of feedback enhances the benefits of retrieval practice (e.g., Butler, Godbole, & Marsh, 2013).

Another issue to consider is the timing of feedback. Instructors can opt to give feedback immediately after tests or even after each question on the test, as is done in intelligent tutoring systems. Alternatively, feedback might come later, such as when instructors give out corrected quizzes a day after they were taken in class. Early evidence suggested that immediate feedback was more useful to learners than delayed feedback (Kulik & Kulik, 1988). More recent research, however, suggests that delaying feedback by even a few seconds (or longer) improves retention relative to immediate feedback (e.g., Butler & Roediger, 2008; Mullet, Butler, Verdin, von Borries, & Marsh, 2014).

One caveat is warranted: this section regarding feedback is most relevant to verbal learning of the type that takes place in classrooms. There is in fact evidence that feedback provided too frequently during motor skill acquisition can be detrimental to transfer and long-term retention of learning (Schmidt, 1991), suggesting that there are limits to the dosage of feedback for some types of retrieval-based training.

Repeated Testing

Flashcards are sometimes used as a study strategy for students. When using flashcards, a student will read a question or prompt on one side of the card, attempt to recall the answer from memory, and then flip the card over to see if she got the answer right or wrong. Thus, flashcards afford a method of repeated retrieval practice with feedback, when used correctly. Unfortunately, students sometimes make the following mistake: the first time they correctly recall an answer, they place that card into a pile of "learned" content. This tactic is sometimes called the "drop" method because they drop cards once they get the answer right once (Karpicke, 2009). In fact, this strategy of recalling something correctly once (or maybe twice) is often built into the advice provided in instructions on how to use flashcards. The logic is that the correct answers show students they know the concept, so they should spend their time on concepts they do not yet know. However, as we have illustrated in this chapter, students often misinterpret current performance for evidence of lasting learning. The evidence to date suggests that multiple correct retrievals improve learning more than a single correct recall, which indicates that the flashcard-carrying student should put the correctly recalled card at the end of her study stack for additional practice.

In an experiment designed to examine whether dropping is, in fact, an efficient method of flashcard use or if it may be detrimental, Karpicke and Roediger (2008) had students learn 40 Swahili–English word pairs (e.g., mashua–boat, lesa–scarf) in one of four schedules of practice. In the study + test schedule, each pair received study and test trials (mashua–boat, then mashua–_________) in four cycles through the list of 40 pairs, regardless of performance on test trials. In the remaining three schedules, a pair was dropped fully or partially from subsequent practice once it was successfully recalled once. In the dropout schedule, a pair was dropped completely from subsequent cycles after one correct recall (no additional practice). In the repeated study only schedule, a pair would continue to be shown in all study trials (mashua–boat), but it was no longer tested after one correct recall. In the repeated retrieval only schedule, a pair would continue to be cued for all test trials (mashua–________), but was not shown as study trials after one correct recall. All four schedules were followed by a test 1 week after practice. The first key finding was that the study + test schedule produced more than twice the rate of correct recall on this delayed test than did the dropout schedule (80% vs. 33%, respectively). The second key finding was dropping study trials but continuing retrieval practice (i.e., the repeated retrieval only schedule) did not harm recall relative to the study + test schedule, but that dropping test trials but continuing study trials (i.e., the repeated study only schedule) caused recall to be on par with the lowly dropout schedule (80% for repeated retrieval without restudying, vs. 36% for repeated studying without retrieval; see Figure 14.2). The take home: retrieval practice after successful recall greatly increased long-term retention, whereas additional studying after successful recall added no benefits to retention. When using flashcards—or any other method that involves retrieval practice—learners should recall information correctly more than once. Karpicke (2009) showed that, when left to their own devices, students often drop flashcards after only one or two correct recalls, but when students were held to a higher criterion of the number of correct recalls, they performed much better on delayed tests.

How many times should one recall something correctly before it has been mastered? Rawson and Dunlosky (2012) provide this advice, based on a combination of spacing (covered in the next section) and retrieval practice with

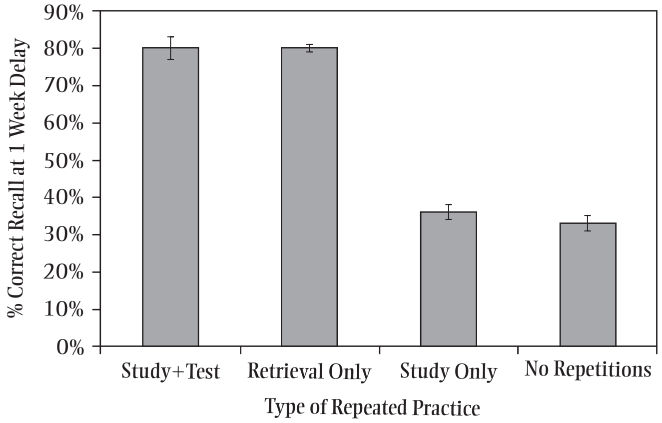

Figure 14.2 The results of Karpicke and Roediger (2008). Following one correct recall, students either continued both study and test trials (far left bar), selectively dropped either study trials (2nd from left) or retrieval trials (3rd from left), or dropped both study and retrieval trials (far right). The two schedules that contained continued retrieval practice (two left bars) produced much better long-term retention than the two schedules that did not incorporate repeated testing (two right bars). Error bars represent standard errors of the means and were recreated from the original figure.
Adapted from J. D. Karpicke & H. L. Roediger, The critical importance of retrieval for learning. *Science, 319,* 966–968, 2008.

feedback: learners should correctly recall content three times during initial acquisition, followed by three additional test trials with feedback spaced out over long periods of time. To keep some fact or concept or procedure at your mental fingertips, you should practice repeatedly at spaced intervals.

MASSED VERSUS SPACED PRACTICE

Learners as well as educators are typically aware that repetition leads to deeper, more lasting learning (Karpicke, Butler, & Roediger, 2009). Unfortunately, that repetition often takes the form of re-reading notes and book passages immediately before a test. In order to achieve the desired long lasting effects, however, learning must be repeated over a much longer period of time. The question that follows, of course, is when should repeated learning be used and how should it be scheduled? Below, we review research on spaced learning and offer guidelines for scheduling repeated learning sessions.

The Spacing Effect

The spacing effect, or the distributed practice effect (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), refers to the finding that repeating study sessions spaced out over time result in better learning than repeated study sessions completed back to back. Spaced practice occurs when the study of a skill or piece of information is practiced with some amount of time between repetitions. Massed practice occurs when study of a skill or piece of information is practiced repeatedly with no time between repetitions. To return to our baseball example, the baseball player who practices hitting a curve ball repeatedly before switching to a fastball is using massed practice. Spacing can be achieved within one practice session (e.g., practicing hitting many different types of pitches each day at practice) or between different practice sessions (e.g., practicing batting on Monday, Wednesday, and Friday while practicing fielding on Tuesday and Thursday). Both within-session and between-session spacing create better learning than massed practice. Within-session spaced practice can be referred to as *interleaved practice* when the time between repetitions is filled by practicing different material. We expand on the benefits of interleaving later in this chapter.

The key to spacing research is that the distribution of time spent studying is manipulated while the total amount of time spent studying is held constant. The classic spacing effect literature involves participants memorizing a list of words to be recalled at a later test. Typically, some words are studied back to back (massed), whereas other words are studied once and then repeated at longer intervals. Then, both groups are tested after some delay. Distributed presentations of words lead to greater recall than massed presentations, and the greater the amount of time (or number of words) between repetitions, the better is later recall (e.g., Melton, 1970). The basic spacing design has been extended into classroom settings as well, with students learning material such as vocabulary definitions (Sobel, Cepeda, & Kapler, 2011) and US history facts (Carpenter, Pashler, & Cepeda, 2009) with the same positive impact.

As noted earlier in this chapter, massed practice typically feels more productive than spaced practice, especially in learning of skills. This may be due to the fact that massed practice allows learners to acquire knowledge more rapidly, even though the long-term retention for the knowledge is poorer (Son & Simon, 2012). Balota, Duchek, and Paullin (1989) found that when students were tested soon after the second study presentation, massed practice led to greater recall than did spaced practice. When tested after a delay, however, the spaced practice condition led to superior recall relative to massed practice. This may be why many students choose to cram for a test just before it. By using massed practice immediately before the test, students can recall the necessary information reasonably well (Roediger & Karpicke, 2006). If asked to recall the information a week (or more) later, however, those students would not have retained much of what they studied relative to spaced practice (or retrieval practice).

Generalization of Concepts

Spacing is effective for more than simply learning and repeating information. Several studies have used a spaced learning design to show improvements in students' abilities to generalize concepts to novel material. In a study by Gluckman, Vlach, and Sandhofer (2014), elementary school students learned about the food chain through lessons covering four different biomes. One group of students received all four biome lessons on the same day (massed condition), while another group learned about one biome each day for 4 days (spaced condition). Total time spent learning was held constant. Students were tested on their ability to recall facts from the lessons as well as on their ability to generalize concepts beyond what was taught in the lessons. Generalization questions were either simple (e.g., bigger animals usually eat smaller animals) or complex (e.g., for a given scenario involving insect population, "What do you think happens to the number of turtles in the swamp?" p. 269). The final test occurred 1 week after the final lesson. Fact questions were based on the biomes taught in the lessons, while the generalization questions were based on a biome that was not covered in the lessons. Thus, for the generalization questions, the students had to apply the principles they had learned to material they had never studied. On the final test, students in the spaced condition performed better on all three types of questions compared to those in the massed condition. They scored 13% higher on the fact questions, 16% higher on the simple generalization questions, and 21% higher on the complex generalization questions. Not only did the spaced schedule improve memory for facts, but it also enhanced the students' ability to transfer what they learned to new contexts.

Similarly, Bird (2011) used spaced practice to teach native Malay speakers English syntax in a college course, using two different levels of spacing of lessons (3 days vs. 14 days). The students practiced reading sentences on a worksheet and identifying errors in syntax. Students received worksheets either 3 days apart or 14 days apart. On a delayed test 60 days after the last study session, the students were asked to identify errors in syntax on novel sentences that had not been included in any previous practice session. Students who practiced identifying syntax errors 14 days apart performed 12% better on the final test of new sentences compared to the group who practiced 3 days apart. This study demonstrates that not only does spaced practice enhance the generalization of knowledge to new contexts, but also that the benefit can last for months following the final practice session. Furthermore, the greater the spacing, the better the long term retention.

Research in Skill Training

Research on the spacing effect has been well established in laboratory and classroom settings with verbal learning, but the benefits of spaced practice can be seen in skill training as well. For example, Moulton et al. (2006) studied medical students as they attempted to learn a difficult microsurgical technique to reattach tiny blood vessels. The lesson consisted of four sessions of instruction, which are usually given to medical students in 1 day. One group of students completed all four sessions in 1 day, as usual, whereas the other group of students completed the same set of lessons with 1 week between each session. When asked to perform the technique on rats 1 month after the fourth lesson, the students who learned in spaced lessons performed better than their peers who learned in the massed fashion. The spaced group took less time to complete the surgery, used fewer hand movements during surgery, and had more success reattaching the vessels. In fact, all the students in the spaced group completed their surgery while 16% of the students in the massed group damaged the vessels beyond repair and could not complete the surgery. Their rats died.

Heidt, Arbuthnott, and Price (2016) showed that spacing is effective in teaching the Enhanced Cognitive Interview (ECI) to police officers as well. The ECI is an interview technique used by the police to questions eyewitnesses with as little interviewer bias as possible. The officers were given 2 hours of training. One group completed the training in 1 day (massed), while the other group completed the training in two sessions of 1 hour each with 1 week between sessions (spaced). Again, the spaced group outperformed their massed counterparts by using more open-ended questions as opposed to leading questions in their interviews.

The Right Amount of Spacing

Up to this point, we have reviewed the evidence that spaced practice results in better long-term learning than massed practice. But how much time should one place between practice sessions? Additional research on spacing has revealed that there is no one spacing interval that always works. For example, in one study students learned face-name pairs until they knew 70% of them. Then they received one final learning episode either 10 minutes or 24 hours later. When they were given a final test 1 day after their last study trial, there was no benefit of spaced presentation. In other conditions, however, students were tested after a week. Then the spacing effect emerged, with those having a spaced presentation after 24 hours recalling more word pairs than those who had only 10 minutes separating original learning from relearning (Pyc, Balota, McDermott, Tully, & Roediger, 2014). Figure 14.3 illustrates this pattern of results.

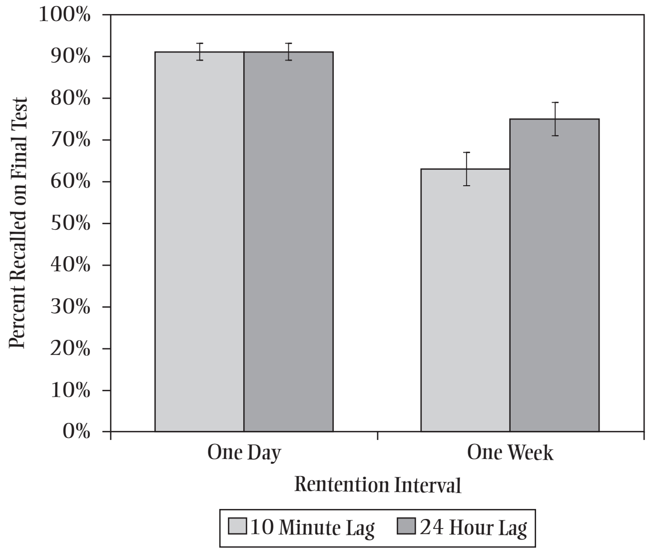

Figure 14.3 The results of Experiment 2b (left pair of bars) and Experiment 4 (right pair of bars) by Pyc, Balota, McDermott, Tully, and Roediger (2014). Participants studied face-name pairs in two separate study sessions either 10 minutes apart or 24 hours apart. When tested 1 day after the second study session, the groups showed no difference in performance. When tested 1 week after the second study session, the group that studied with 24 hours between sessions performed better than the group that studied with only 10 minutes between sessions. Error bars represent standard errors of the means and were recreated from the original figure.
Adapted from M. A. Pyc, D. A. Balota, K. B. McDermott, T. Tully, & H. L. Roediger, Between-list lag effects in recall depend on retention interval, *Memory & Cognition, 42*, 965–977, 2014, copyright 2014 by Psychonomic Society, Inc. Reprinted with permission of Springer.

The ideal interval for spacing likely depends on many different factors, but the factor identified in the Pyc et al. study (2014)—the retention interval between study and test—has been shown repeatedly to affect the success of spaced practice. Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) provided a broad assessment of various spacing gaps and retention intervals. More than 1,000 people studied obscure trivia items in two sessions (initial study and review) with nine different spacing intervals ranging from 0 days (the massed condition) to spacing of up to 105 days between sessions. Participants were tested on the trivia 7, 35, 70, or 350 days after their review session. All the groups given spaced presentation outperformed the massed practice group. The most successful groups scored up to 64% higher than the massed practice group. But there was no best spacing interval for all conditions; rather each retention interval required a different spacing interval for best performance. For example, for those tested 350 days after their review session, the group with 21 days between sessions performed the best. For those tested 1 week after their last review session, the group that used 1 day of spacing performed the best. These results show that there is a balance to be struck between the spacing of practice sessions and the time until the final test. Using the longest spacing interval between practice sessions isn't always the best route to take, if the final assessment will occur relatively quickly after learning. Further research is required to determine the optimum spacing schedule for various retention intervals, but the research completed so far allows us to make educated estimates. After reviewing the available literature, Putnam, Nestojko, and Roediger (2016) suggested the following guideline: 1 day of spacing should be used for 1 week of retention, 1 week of spacing should be used for 2 months of retention, and 1 month of spacing should be used for 1 year of retention.

In general, the longer you would like to remember a topic or skill, the more spacing should occur between practice sessions. Practice sessions should not, however, be so far apart that forgetting effectively erases the benefits of the previous session. Often, spacing research involves simply presenting the information again. That works to a degree. However, if retrieval attempts are made before restudy in spaced conditions, the beneficial effect of spacing repetition is much greater (Agarwal, Finley, Rose, & Roediger, 2016). And, as noted in the previous section, to maintain knowledge for the long term, repeated practice sessions at widely spaced intervals are important. For example, if medical students know that a test of surgical techniques will occur at the end of the year (and will be needed in their practice), they are better off practicing those techniques once a month for the entire year than massing their study when the techniques are first taught.

Spacing requires planning, and such planning becomes increasingly important when one needs to complete multiple courses or learn multiple skills within the same time frame. In those cases, it becomes important to overlap spacing schedules. One might practice seven different skills each week, devoting 1 day to practicing each skill. That way, each skill is practiced with a 1-week spacing interval, but no days are wasted waiting for the next practice session. How time is spent during each practice session is equally important. Earlier in this section, we briefly mentioned a type of within-session spacing known as interleaving. In the following section, we will review the interleaving literature and provide examples of how an interleaved practice schedule can benefit learning in several different domains.

Spacing Suggestions for Operational Settings

Most training in all settings is massed, for reasons discussed earlier. Massed training leads to quick gains in learning and both students and trainers tend to favor it for this reason. Massed training, however, leads to rapid decay of knowledge and skills. The trainers of Navy SEALS worried about the massed nature of their training because training of critical skills (e.g., parachute jumping) is taught in a massed fashion. In the case of learning this skill, spacing of training after initial learning is difficult because training must involve coordination with other units that fly the airplanes (and getting the airplanes themselves to the SEAL base in Coronado, California). Thus, practical considerations can limit spaced training.

Because military training is composed of weeks (or more) of learning difficult skills in many different skill sets, spacing and interleaving should be worked into the training schedule whenever possible. In fact, there may already be areas of training that use spacing techniques. Military training incorporates huge amounts of repetition to instill necessary skills in the soldiers. Spacing those repetitions over time rather than running them back-to-back is a simple and straightforward way of employing the spacing effect. Drills that a soldier is expected to know by heart should be revisited every week or perhaps every month in order to prevent forgetting. The Jumpmaster school in the US Army, for example, is already using the principle of spacing by requiring their graduates to attend meetings and practice their skills periodically even after the formal classes are finished (Kienery & Lahr, personal communication, 2016). In addition to practicing drills, spacing can be employed in classroom settings. Revisiting important points each week in a class or having regular quizzes over recurring themes in the curriculum will help to reap the benefits of both the spacing effect and the retrieval practice effect. Taking the amount of instruction and practice that is already being used and rescheduling it so that consolidation can occur between lessons can go a long way to improve the efficiency and efficacy of existing military training programs.

One issue encountered in practical situations is when to introduce spaced practice. After all, some basic learning must occur before a trainee can practice at spaced intervals. This is particularly true for those skills that can result in severe injury or death when performed incorrectly. Although there is not much evidence yet to back up this recommendation, one practical solution is to have a bit of massed practice to gain a certain level of knowledge about each skill to be learned and then to begin to space out the sessions of practice on those skills.

INTERLEAVED PRACTICE

In our section on the spacing effect, we discussed the importance of organizing study and practice sessions over time. In this section, we review the research pertaining to the most effective ways to arrange individual practice examples within sessions. Mixing many different types of examples (interleaving) is far more effective for long-term learning than grouping together similar examples (blocking) within a practice session. The following section will provide guidance on how to design each practice session in order to achieve long-term retention.

Interleaving Versus Blocking

Let's revisit the baseball player example once more. On the days he practices batting, he needs to review curve balls, fastballs, and change ups (and possibly other pitches, but we'll stick with three for now). He can begin practice by hitting a curve ball many times in a row. Then he can move on to the fastball, then the change up. This type of practice schedule is referred to as blocking (a particular type of massed practice). Alternatively, he can mix up hitting all three pitches throughout the session. This type of practice schedule is referred to as interleaving. There is strong evidence that an interleaved practice schedule results in better long-term and flexible learning in at least three domains: motor skill learning, category learning, and mathematical problem-solving.

Much like in spacing research, a key principle of interleaving research is that the number of practice trials in a study session remains constant across groups. Only the organization of those trials is manipulated. The classic interleaving experiment typically involves asking participants to learn a set of related skills in a specific order, then testing those skills after some delay. Rohrer and Taylor (2007), for example, asked students to learn the proper equations used to calculate the volume of four solid three-dimensional shapes. One group of students followed a blocked schedule in which they practiced calculating the volume of one type of solid per session (much as occurs in classrooms). Another group of students followed an interleaved schedule in which they practiced calculating the volume of all four types of solids during every practice session. There were four practice sessions total, spaced 1 week apart. Both groups were then tested 1 week after the fourth practice session. During initial learning, the blocked group reached a higher level of performance than did the interleaved group. On the final test, however, the interleaved group calculated the volumes of the solids correctly 63% of the time, whereas the blocked group was only able to calculate the volumes correctly 20% of the time.

With such a dramatic difference in performance (43%!), why isn't interleaving a more widely used technique? As in our discussion of massed practice, blocked practice sessions feel more productive (Rohrer & Taylor, 2007) because initial learning is faster if the learning is blocked. In the experiment just described, students in the blocked group calculated the volumes of the solids more successfully than their interleaved peers *during* each practice session (89% compared with 60% accuracy overall). They learned the calculation more quickly and implemented them successfully, but they did not retain the knowledge after a week's delay. Why? Because with blocked practice, students know the type of volume they have and they know the formula for calculating its volume. So they just apply the one formula over and over. In the interleaved learning situation, however, they must determine what type of solid is being described, and they must then retrieve the appropriate equation for that volume. The final test on all four volumes requires the skill of determining what type of solid is being described and retrieving that equation. Students who learned by interleaving practiced this skill, whereas those learning through blocked practice never had to learn this discrimination. Thus, using a term from earlier in the chapter, interleaved learning leads to greater transfer-appropriate processing than does blocked learning because tests almost always require students to figure out what type of challenge they are facing.

Other interleaving experiments have obtained similar results. Kornell and Bjork (2008) had students learn some material on a blocked schedule and other material on an interleaved schedule so that all students experienced both types of practice sessions. After the final test, they asked students which schedule helped them to learn better. Even though test scores were much better for the material learned on an interleaved schedule, 78% of the students reported that they learned the blocked material as well as or better than the interleaved material. That is probably because blocked learning seems easier than interleaved learning.

In another study, a group of undergraduates were presented with an explanation of both blocked and interleaved study schedules. When asked to choose which would lead to better learning, more than 90% chose the blocked schedule (McCabe, 2011). Whether the reason for this preference is due to the typical structure of lesson plans and textbooks or due to the feeling of success achieved during blocked practice is unknown. Another possible reason is that interleaved learning feels hard and leads to more errors (a desirable difficulty). We will see that interleaving is also more successful in the learning of motor skills.

Research in Motor Skill Training

For decades, we have known that interleaved practice produces better long-term retention for motor skill learning. The first study to demonstrate the effect occurred nearly 40 years ago using patterned arm movements (Shea & Morgan, 1979). Since then, the literature has expanded to include studies in the domain of sports training, playing musical instruments, and surgical training. In each field of skill acquisition, the same patterns are observed: blocked practice leads to more rapid skill learning, but interleaved practice leads to better long-term retention and transfer to novel situations.

In a study by Goode and Magill (1986), college students with no prior experience were taught three types of badminton serves on either a blocked or interleaved practice schedule. For both schedules, students practiced 3 times per week for 3 weeks (9 sessions total). The blocked group practiced 36 short serves on one day, 36 long serves the next day, and 36 drive serves the following day. This schedule was repeated each week. The interleaved group practiced 12 serves of each type (short, long, and drive) in random order during every session. For both groups, the serves were practiced only on the right side of the court. One day after the last practice section, all students were tested on each of the three serves they had learned. Compared to the blocked group, the interleaved group was able to land more accurate serves not only from the right side of the court (where they had practiced) but also from the left side of the court (from which they had never served during the experiment). This study illustrates that interleaved practice leads to improved motor skill learning both in contexts that match the practice conditions and in new contexts where the learner must adjust their skills to the environment. Similar results have been found in studies on baseball batting (Hall, Domingues, & Cavazos, 1994), golf (Porter, Landin, Hebert, & Baum, 2007), and volleyball (Kalkhoran & Shariati, 2012).

The advantages of interleaved practice are not limited to novices. A study of formally trained piano players by Abushanab and Bishara (2013) demonstrated the same pattern of results. In this study, rather than learning how to play the piano, pianists learned to play new melodies with the goal of increasing their speed and accuracy on each melody. Each pianist practiced some melodies using a blocked schedule and other melodies using an interleaved schedule. During the practice session, the pianists were able to reach higher speeds on the melodies practiced on a blocked schedule compared to the interleaved schedule. When tested on the melodies 2 days later, however, melodies practiced using the interleaved schedule were played more quickly than those practiced using the blocked schedule. Accuracy was high for both groups during practice and test due to the fact that all the participants were experienced piano players. Unlike studies that typically use novices, this study illustrates that learners with prior experience can use interleaved practice to hone their existing skill sets. In military training, for example, each soldier needs to know what to do if the firing pin on their rifle gets jammed. By interleaving various rifle maintenance techniques during training, even experienced soldiers can shave precious seconds off the time required to recall the appropriate action and repair their rifle during a combat situation, where each second is critical.

Recently, Welsher and Grierson (2017) used interleaving to study a training task used to teach medical students specific hand movements within a confined space meant to imitate the conditions of a laparoscopic surgical procedure. The students had to pick up a bean with a surgical tool using their dominant hand, pass it to their nondominant hand, and then deposit the bean into a dish. The task was confined to a small box, which restricted the hand movements that the students could use to complete the task. In this study, medical students observed experts and novices as they completed the training task. The students also practiced the task themselves intermittently between observations. One group's observation was scheduled using a blocked format, while another group's observations were scheduled using an interleaved format. In the blocked schedule, students observed an expert completing the training task 20 times back to back and also observed a novice (one of their peers) completing the task 20 times back to back. In the interleaved schedule, the expert and novice alternated completing the training task until each had performed the task 20 times. For both groups, students attempted the task themselves after each set of 10 observations. At the end of the final observation, the students were tested on their ability to complete the training task. The following day, the students were tested on their ability to complete that task again by leading with their nondominant hand (all practice trials and tests had been lead with the dominant hand).

Interestingly, performance by both groups improved at the same rate during the observation and practice trials, and performance on the immediate test was also nearly identical for both groups. Differences between the groups emerged during the test using the nondominant hand. When leading with their nondominant hand, the blocked group produced more than 50% more errors than the interleaved group. The blocked schedule was far less successful when the students had to adjust their skills to a new procedure, even one that closely resembled what they had been practicing. Thus, the interleaved observation served to make learning more flexible and applicable to a new situation.

Research in Category Learning

Interleaving examples offer benefits to differentiating between categories within a group as well. Kornell and Bjork (2008), discussed earlier in this section, performed a well-known study of the interleaving advantage in category learning using painting styles. In this study, students were asked to learn the styles of 12 different artists by viewing 6 paintings done by each artist. For 6 of the artists, paintings appeared in a blocked format (all 6 paintings by one artist in a row). For the other 6 artists, paintings appeared in an interleaved format (no 2 paintings by the same artist were shown back to back). Fifteen minutes after the last painting was studied, students were shown new paintings (4 for each artist in random order) and asked to determine which artist created each painting. Students were better able to correctly identify the artist of the novel paintings when the work of that artist had originally been studied in the interleaved condition relative to the blocked condition (although students believed blocked learning led to better performance).

Kang and Pashler (2012) sought to determine why the interleaved schedule resulted in better category learning than the blocked schedule in the Kornell and Bjork (2008) experiments. In one of their new experiments, paintings by different artists were presented simultaneously (side by side) in addition to the interleaved and blocked schedules. They found that the simultaneous presentation resulted in learning that was just as good as the interleaved learning schedule (both were superior to blocked learning). This result suggests that the enhanced learning in the interleaved condition may be a product of students contrasting the styles of different artists. They learned what features differ between various styles more easily than in the blocked case. In the blocked case, people probably concentrate on the commonalities between different works by the same artist.

Interleaving benefits have been found in cases of auditory discrimination as well. Chen, Grierson, and Norman (2015) conducted a study in which nursing students were taught to discriminate between different respiratory and cardiac diagnoses by listening to internal body sounds through a stethoscope. The nursing students listened to recordings of 8 different diagnoses (4 respiratory disorders and 4 cardiac disorders). The students heard three example of each diagnosis using either a blocked or interleaved schedule. Immediately after studying the recordings, the students were tested on their ability to match the correct diagnosis with the recording. Half of the recordings on the test were taken directly from the study session while the other half were new recordings of the diagnoses they had studied. The students also took the same test 1 week later. Nursing students in the interleaved group outperformed the blocked group for both respiratory and cardiac disorder identification on both the immediate and delayed tests. The difference in performance was particularly dramatic for the new recordings. As you can see in Figure 14.4, the interleaved group scored somewhat higher than the blocked group on the recordings they had heard during study. For the new recordings, however, the difference in performance is much more striking. In fact, the interleaved group performed just as well on the new examples as the blocked group did on examples they had heard before. The field of medicine is rich with studies demonstrating the advantage of interleaved practice in category discrimination. Similar patterns of results have been found in psychiatric diagnoses (Zulkiply, McLean, Burt, & Bath, 2012) and electrocardiogram reading (Hatala, Brooks, & Norman, 2003), as well.

Research in Mathematical Problem-Solving

Doug Rohrer's work on interleaved practice in mathematical problem-solving makes up a substantial portion of the interleaving literature. We described one study earlier, the one involving solving for volumes of different kinds of solids. Through laboratory and classroom research, Rohrer and his colleagues have been able to dissect the interleaving advantage in order to understand the underlying patterns of performance within the field of mathematics. In one study, for example, Taylor and

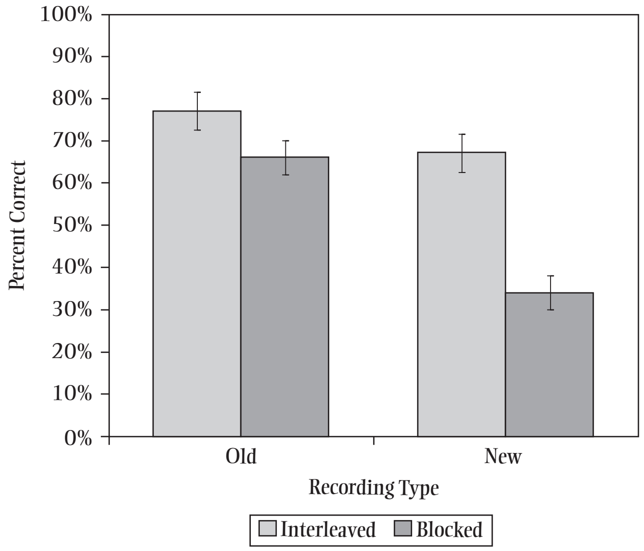

Figure 14.4 The results of Experiment 3 by Chen, Grierson, and Norman (2015). Nursing students tested on respiratory and cardiac disorder recordings performed better on immediate and delayed tests when study followed an interleaved schedule compared to a blocked schedule. When test performance was divided between old (previously studied) and new (previously unstudied) recordings (shown here), there was a greater difference in performance for new recordings than for old recordings.
Adapted from R. Chen, L. Grierson, & G. Norman. Manipulation of cognitive load variables and impact on auscultation test performance, *Advances in Health Sciences Education, 20*, 935–952, 2015, copyright 2014 by Springer Science+Business Media Dordrecht. Reprinted with permission of Springer.

Rohrer (2010) asked elementary school students to solve math problems involving prisms. Students had to determine the number of faces, corners, edges, and angles on various prisms. Each type of problem could be solved using a different formula, which the students learned how to apply. Students completed a series of practice problems in either a blocked or interleaved format. The following day the students took two tests. For both tests, all of the math problems were new cases of the types of problems they had studied. That is, none of the practice problems was included on either test. On the first test, no formulas were provided. Students had to recall and apply the appropriate formula for each problem. On the second test, the formula for each type of problem (face, corner, edge, and angle) was provided at the top of the page.

During the practice session, students in the blocked group performed much better than their classmates in the interleaved group. The blocked group got nearly every practice problem correct, while the interleaved group only got between 70% and 80% correct. On the first test, however, the interleaved group performed better than the blocked group. Interestingly, the interleaved group's performance stayed at roughly the same level it had been at during practice, 77%. The blocked group's performance suffered severely, dropping from near perfect during practice down to 38% on the test. On the second test, in which students were given the formulas and told which formula applied to which problem (e.g., "faces = b + 2," p. 842), the groups showed nearly no difference in performance. The blocked group returned to near perfect performance (90%) and the interleaved group achieved perfect performance (100%). This change in performance could either be because the blocked group couldn't remember the formulas they learned or because they were using the right formulas but on the wrong test problems.

Taylor and Rohrer (2010) addressed this issue by analyzing the types of errors made on the first delayed test. For both the interleaved and blocked groups they identified two types of errors: fabrication errors and discrimination errors. *Fabrication errors* were those in which the student used a formula that was not taught during the experiment. Fabrication errors included misremembered formulas and formulas learned in other units of the math class. Making a fabrication error would suggest that the student simply couldn't remember the correct formula. *Discrimination errors* were those in which the student used a formula taught during the experiment, but used it on the wrong type of problem (e.g., using the faces formula to calculate the number of edges on a prism). Making a discrimination error would suggest that the student couldn't remember how to apply the formulas they learned. Remarkably, the pattern of errors differed dramatically between the blocked and interleaved groups (see Figure 14.5). The interleaved group's errors were evenly split between fabrication and discrimination errors. The blocked group, however, made three times as many discrimination errors as fabrication errors. Additionally, the students in both groups made about the same number of fabrication errors. So, the students in the blocked group weren't having any more trouble remembering the formulas than their classmates in the interleaved group. Rather, the blocked group was having trouble knowing when to apply the formulas they had learned.

In his chapter on interleaving for classroom learning, Kang (2016) perfectly sums up the advantage of interleaved practice, "it is not sufficient to learn *how* to execute a strategy; one must also know *when* a particular strategy is appropriate" (p. 86). In each study described in this chapter, we discussed how interleaved practice led to a better ability to apply the skills learned to novel situations. Since no training program or classroom lesson can cover every instance where a skill might be needed, transfer of the skill to new contexts is crucial. For this reason, we endorse interleaved practice over blocked practice wherever it can reasonably be applied. Learning will be slower and more effortful, but will result in more long lasting, flexible understanding.

Interleaving Suggestions for Military Training

Interleaving has not yet been adopted in many instructional fields. Classrooms across the country teach their curriculums using subject units and assign homework in a blocked format. Similarly, in virtually all arenas of training of which we are aware (sports, music, business, and the military), blocked training is the norm. Yet, as we have described, the benefits of interleaving have been found in a diverse range of settings. Studies in education from elementary school and medical school, in sports, music, and art all benefit from interleaving. Military training is no different.

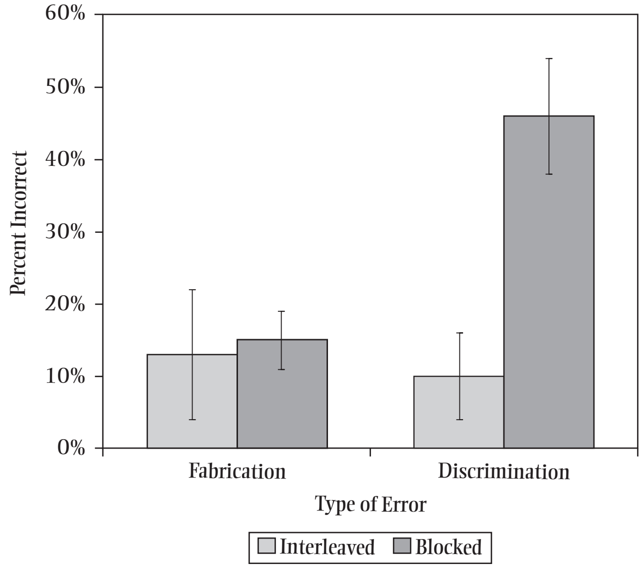

Figure 14.5 Errors made on the first test in the study by Taylor and Rohrer (2010), expressed as percent incorrect for two types of errors (fabrication and discrimination), plotted as a function of practice condition (interleaved vs. blocked). In this experiment, elementary school students learned formulas to solve four different types of math problems about prisms. Students practiced using the formulas one problem type per practice set (blocked) or all four problem types in each practice set (interleaved). They were tested the following day on a set of new problems that required them to recall and use the formulas they had learned the previous day. The interleaved group produced fewer errors over all (the sum of the two white bars), with significantly fewer discrimination errors than the blocked group. Error bars represent standard errors of the means and were recreated from the original figure.
Adapted from K. Taylor & D. Rohrer, The effects of interleaving practice. *Applied Cognitive Psychology, 24,* 837–848, 2010, copyright 2009 by John Wiley & Sons, Ltd. Reprinted with permission of John Wiley and Sons.

For example, a soldier in combat, and especially an officer in charge of directing soldiers, must know at a moment's notice which tactical formation is best for the given situation. Training programs cannot possibly be expected to run drills on every scenario that might occur in a combat setting; time is too limited and combat is too unpredictable. Instead, a training program should prepare soldiers to apply the principles they learn to novel situations, which is exactly what an interleaved practice schedule does. Much like spacing, interleaving can be achieved simply by rearranging the existing practice drills so that soldiers learn not only what to do in a particular formation, but also when that formation is more useful than another.

Interleaving can be useful outside of learning tactical formations as well. According to the US Army's *Soldier's Manual of Common Tasks: Warrior Skills Level 1* (2015), each army soldier must learn to "identify terrain features on a map" (p. 3–49). This type of learning is perfectly suited to interleaved practice, specifically as discussed in our section on category learning. A soldier could implement an interleaved practice schedule during his own individual study. By studying maps with different types of terrain in random order, the soldier would be able to differentiate between hills, saddles, and valleys (for example) more quickly and efficiently than a soldier who studied the same maps in a blocked schedule. Using the interleaved format would also result in a soldier who is able to quickly and accurately read a map that he or she has never seen before. Implementing interleaving across the various facets of military training will produce soldiers who can apply their diverse knowledge to any situation effectively.

CONCLUSION

To summarize, retrieval practice, spaced practice, and interleaved practice are each useful, scientifically backed methods to create long-term, flexible learning in a variety of settings with many different types of materials. Our review of the literature in this chapter has summarized research in laboratory, classroom, and professional settings. We have shown that the techniques we suggest are effective on materials as simple as lists of words and as complex as microsurgical procedures. It should come as no shock then that the best recommendation we can make to improve military training programs is to implement spaced, interleaved retrieval practice wherever possible. And although we have focused our attention on how these techniques relate to military training programs, they certainly apply in training personnel in other high-stakes settings. Agencies that train first responders—such as police officers, firefighters, and emergency medical technicians (EMTs)—will benefit from these techniques. We will close this chapter with a few recommendations on how to apply each of these principles.

Putting It All Together: Advice for Implementing These Principles

Retrieval Practice

Retrieval practice is already being used in military training extensively for skills training. For example, soldiers learn how to fire a weapon accurately and consistently by practicing often on firing ranges designed to mimic combat situations. Retrieval practice can be used in classroom settings as well by adding frequent, low-stakes quizzing to the existing curriculum. In this way, soldiers are required to recall and apply their training in every stage of learning throughout the program. Additionally, soldiers should be provided with feedback for incorrect answers or for errors, allowing them to identify and improve on weak areas for future tests. Finally, quizzes (as well as drill practice, weapons training, etc.) should be repeated frequently throughout the program. Retrieving knowledge from memory once is insufficient for long-term retention. A soldier should be required to retrieve the knowledge gained from his training repeatedly and throughout the course of training.

Spaced Practice

We discussed the benefits of spaced learning over massed learning at length in this chapter. One simple way to use spacing is to take a lesson that is typically taught in a day-long seminar and break it up into several smaller lessons spaced out a week apart. Reviews of the learned material (ideally through retrieval practice) should also be spaced out over time to prevent forgetting. The spacing schedule for any individual training program depends on the retention interval (the time until the final test). Longer retention intervals require longer spacing between sessions. In general, 1 day of spacing is useful for 1 week of retention, 1 week of spacing is useful for 2 months of retention, and 1 month of spacing is useful for 1 year of retention. Soldiers need training to last throughout their time in the military, so continual spaced refreshers for practicing critical skills should be implemented. A spacing schedule can expand as time goes on, to reflect the changing needs of the soldiers. Perhaps 1 day of spacing is needed while soldiers initially learn a new tactical formation. After they learn the basics, spacing can be increased to one training session per week, to find and correct weak areas or common errors. Finally, when the soldiers can routinely perform the formation correctly, spacing is increased to once per month, so that forgetting over time is minimized.

Interleaved Practice

Interleaving research has shown that simply mixing up the order in which skills are practiced can have dramatic effects on retention and flexibility of learning. A soldier who must learn to repair a machine or vehicle in the field, for example, would be better off if he or she practiced fixing many different mechanical issues in random order rather than practicing the same one over and over before moving on to the next issue. In the field, soldiers will rarely be told what is wrong before they start working. They will need to assess the situation to figure out which solution to apply and then apply that solution correctly, potentially under stressful conditions and time restraints. By mixing up practice during training, soldiers gets the opportunity to learn what different issues look like and how to tell one issue from another, as is evidenced in the category learning literature. A soldier will also more easily remember which solutions to use in which situations, as discussed in the literature on mathematical problem-solving. Even if the soldier encounters a problem in the field that she had never seen during training, the soldier who practiced using an interleaved schedule would be best equipped to apply the principles she learned to the new situation successfully.

Automated Training Using These Principles

A major challenge in training is how to get thousands of people up to speed without spending too many resources. There is a long tradition of the military utilizing psychological tools and services en masse, from the intelligence tests and personnel tests used for recruitment and placement during World War I to the clinical treatment of soldiers returning from World War II (Fancher & Rutherford, 2012; Hilgard, 1987). A more recent development in this tradition is intelligent tutoring systems designed to train soldiers. For example, the Immersive Naval Officer Training System (INOTS) is an immersive environment for role-playing with virtual humans designed to train officers' interpersonal skills (Hays et al., 2012). At the same time the military has been developing virtual training environments for various skills, the education industry has been designing products aimed at college students and instructors to optimize learning in and out of the classroom (e.g., Assessment and LEarning in Knowledge Spaces, ALEKs; Falmagne, Cosyn, Doignon, & Thiéry, 2006).

We propose a convergence of these simultaneous developments. Specifically, we believe the military is uniquely situated to develop adaptive learning programs that utilize the cognitive principles outlined in this chapter. Such a learning system could provide individual soldiers with training tailored to their performance to optimize long-term retention and transfer of skills by implementing schedules of spaced, interleaved retrieval practice. Furthermore, this type of digital learning environment would be cost-efficient in the long run, given that classroom training costs time and money. Our idea is not, of course, completely novel. For example, the Air Force Research Laboratory has engaged with cognitive psychologists to develop and assess learning algorithms for intelligent tutoring systems (Jastrzembski, Gluck, & Gunzelmann, 2006). Our goal in discussing intelligent tutoring systems in this chapter is to encourage further development of these systems by the military, specifically focusing on implementing spacing, interleaving, and retrieval practice. If you want thousands of soldiers or emergency personnel to efficiently learn skills that will last, designing any learning environment—digital or traditional classroom and field training—with these principles in mind will certainly help.

REFERENCES

Abushanab, B., & Bishara, A. J. (2013). Memory and metacognition for piano melodies: Illusory advantages of fixed-over random-order practice. *Memory & Cognition*, *41*, 928–937.

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701.

Agarwal, P., Finley, J. R., Rose, N., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory, 25*(6), 764–771.

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging*, *4*, 3–9.

Bird, S. (2011). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *32*, 437–442.

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

Brown, P. C., Roediger, H. L, & McDaniel, M. A. (2014). *Make It Stick: The Science of Successful Learning*. Cambridge, MA: Harvard University Press.

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

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*, 290–298.

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

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

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology*, *23*, 760–771.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19,* 1095–1102.

Chen, R., Grierson, L., & Norman, G. (2015). Manipulation of cognitive load variables and impact on auscultation test performance. *Advances in Health Sciences Education, 20,* 935–952.

Couch, D. (2004). *The Finishing School.* New York: Three Rivers Press.

Couch, D. (2009). *The Warrior Elite: The Forging of SEAL Class 228.* New York: Three Rivers Press.

Falmagne, J. C., Cosyn, E., Doignon, J. P, & Thiéry, N. (2006). The assessment of knowledge, in theory and in practice. In R. Missaoui & J. Schmid (Eds.), *Formal concept analysis* (pp. 61–79). Berlin: Springer.

Fancher, R., & Rutherford, A. (2012). *Pioneers of Psychology: A History*. New York: W. W. Norton & Company.

Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory, 18,* 335–350.

Gagnon, S. A., & Wagner, A. D. (2016). Acute stress and episodic memory retrieval: Neurobiological mechanisms and behavioral consequences. *Annals of the New York Academy of Sciences, 1369,* 55–75.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 40,* 1–124.

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology*, *28*, 266–273.

Goode, S., & Magill, R. A. (1986). Contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport, 57,* 308–314.

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills, 78,* 835–841.

Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education, 8,* 17–26.

Hays, M.J., Campbell, J.C., Timmer, M.A., Poore, J. C., Webb, A. K., Stark, C. & King, T. K. (2012, December). *Can role-play with virtual humans teach interpersonal skills?* Proceedings of the 34th Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Heidt, C. T., Arbuthnott, K. D., & Price, H. L. (2016). The effects of distributed learning on enhanced cognitive interview training. *Psychiatry, Psychology, and Law*, *23*, 47–61.

Hilgard, E. R. (1987). *Psychology in America: A Historical Survey*. New York: Harcourt Brace Jovanovich.

Jastrzembski, T. S., Gluck, K. A., & Gunzelmann, G. (2006). Knowledge tracing and prediction of future trainee performance. In *The Proceedings of the Interservice/Industry Training, Simulation, and Education Conference.* Orlando, FL: National Training Systems Association (pp. 1498–1508).

Kalkhoran, A. F., & Shariati, A. (2012). The effects of contextual interference on learning volleyball motor skills. *Journal of Physical Education and Sport, 12,* 550–556.

Kang, S. H. K. (2016). The benefits of interleaved practice for learning. In J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds.), *From the Laboratory to the Classroom: Translating the Science of Learning for Teachers.* (pp. 79–93). Oxford: Routledge.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97–103.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17,* 471–479.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 13,* 585–592.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58,* 79–97.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition, 39,* 462–476.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20,* 516–522.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596–606.

Moulton, C.-A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect? A randomized, controlled trial. *Annals of Surgery, 244,* 400–409.

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition, 3,* 222–229.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3–8.

Porter, J., Landin, D., Hebert, E., & Baum, B. (2007). The effects of three levels of contextual interference on performance outcomes and movement patterns in golf skills. *International Journal of Sports Science and Coaching, 2,* 243–255.

Putnam, A. L., Nestojko, J. F., & Roediger, H. L. (2016). Improving student learning: Two strategies to make it stick. In J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds.), *From the Laboratory to the Classroom: Translating the Science of Learning for Teachers.* (pp. 94–121). Oxford: Routledge.

Pyc, M. A., Balota, D. A., McDermott, K. B., Tully, T., & Roediger, H. L. (2014). Between-list lag effects in recall depend on retention interval. *Memory & Cognition, 42,* 965–977.

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving durability and efficiency of student learning? *Educational Psychology Review, 24,* 419–435.

Roediger, H. L., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels of processing framework. *Memory, 10,* 319–332.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning. *Psychological Science, 17,* 249–255.

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1155–1159.

Roediger, H. L., Putnam, A. L., & Smith, M. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 55* (pp. 1–36). Oxford: Elsevier.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems improves learning. *Instructional Science, 35,* 481–498.

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

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432–1463.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in Motor Neuroscience* (pp. 59–75). Dordrecht, The Netherlands: Kluwer.

Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching, 16*(2), 179–196.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 179–187.

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science, 354,* 1046–1048.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning, *Applied Cognitive Psychology, 25,* 763–767.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10,* 176–199.

Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review, 24,* 379–399.

Taylor, K., & Rohrer, D. (2010). The effects of interleaving practice. *Applied Cognitive Psychology, 24,* 837–848.

US Department of the Army. (2015). Identify terrain features on a map. In *Soldier's Manual of Common Tasks: Warrior Skills Level 1* (pp. 3-49–3-57). Washington, DC: US Department of the Army.

Welsher, A., & Grierson, L. E. (2017). Enhancing technical skill learning through interleaved mixed-model observational practice. *Advances in Health Sciences Education, 22*(5), 1201–1211.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction, 22,* 215–221.