6

IMPROVING STUDENT LEARNING

Two Strategies to Make It Stick

Adam L. Putnam, John F. Nestojko, and Henry L. Roediger III

WASHINGTON UNIVERSITY IN ST. LOUIS

Introduction

The aim of this book is to take research from the laboratory to the classroom. This is an important goal because, despite decades of research by cognitive psychologists, there is not a strong tradition of translational educational research, where findings are clarified in controlled laboratory settings and slowly introduced to real-world classrooms (Roediger, 2013). Several books written for non-specialists, such as *Make It Stick: The Science of Successful Learning* (Brown *et al.*, 2014) and *Why Don't Students Like School?* (Willingham, 2009), highlight some main points about educationally relevant research. Of course, we hope that books like the one you are holding will also help to spread the word about how findings from cognitive psychology can be used to improve education. In this chapter we take a close look at two strategies—retrieval practice and spaced learning—that most cognitive psychologists agree are some of the strongest candidates in terms of having a significant impact on education.

Students' Understanding of Learning

One question that needs to be posed at the outset is why learning is so hard. After all, humans are natural learners, and we learn from birth without necessarily trying hard to do so. Children do not have to *try* to understand and speak the language that surrounds them when they are growing up—it happens naturally. Why can't they learn to read or do arithmetic the same way? The answer to this question is complex, of course, but let us supplant it with a seemingly easier one. Because children *do* have to learn topics for which they are not naturally prepared (as they are with language), why don't they discover good learning strategies and stick with those? Why is education so hard for so many students?

Students display wide individual differences, of course, but even in excellent universities (and medical schools), students report that they do not know how to study effectively. In addition, when they think they are using good strategies, they may be wrong. Survey studies that have examined what students actually do when they study typically reveal that they choose to read and reread as their primary strategy (e.g. Karpicke *et al.*, 2009). They read the text and highlight or underline certain passages. They then reread the material (especially the critical highlighted parts) repeatedly to prepare for the test. As we shall see later in this chapter, rereading, despite its ubiquity as a study strategy, is not particularly effective. Much better strategies are possible (such as retrieval practice, discussed below). And even when students do reread material, they often do so under less than optimal conditions. That is, they often read materials over and over again with little time between rereadings. One main theme of this chapter is the spacing out of repeated rereadings to make them effective.

Repeated reading seems to have another drawback—it makes students overconfident about what they know (e.g. Roediger & Karpicke, 2006a). If you read something repeatedly, it will become quite familiar and you may be fooled into thinking you understand it when you do not. The familiarity or fluency of the material can be mistaken for knowledge and understanding. Unless the student can actively use the material—that is, they can call it up when necessary to answer questions or solve problems—the information is not useful. The first technique we shall discuss in this chapter, namely retrieval practice via testing, permits students to do this. Testing oneself also helps to correct overconfidence from rereading, because students can learn what they know and what they do not know by quizzing themselves.

The remainder of this chapter is devoted to discussion of the two topics introduced in this section—retrieval practice via quizzing, and spacing. These are certainly not the only topics important to education that have emerged from research in cognitive psychology, but they represent two central principles. Other important points are discussed elsewhere in this volume.

Retrieval Practice

One of the most effective learning tools available to teachers and students is retrieval practice. Also called the testing effect, retrieval practice refers to the idea that retrieving something from memory not only measures what someone has learned, but also changes the retrieved memory, making it easier to recall in the future. Psychologists have periodically studied this concept over the past 100 years (e.g. Abbot, 1909), but the last decade has seen a surge of interest, with researchers exploring both why retrieval practice can enhance memory (e.g. Carpenter, 2009; Karpicke *et al.*, 2014; Pyc & Rawson, 2009, 2010) and how it can be used to improve education (e.g. Karpicke & Blunt, 2011; Karpicke & Grimaldi, 2012; Roediger *et al.*, 2011a) (for reviews, see Roediger *et al.*, 2010; Roediger & Karpicke, 2006b).

Taking a test generally enhances later retention, because tests require some form of retrieval from memory. Even thinking about information and not saying it or writing it down improves retention (e.g. Putnam & Roediger, 2013; Smith *et al.*, 2013). Although the exact mechanisms involved remain unclear, retrieving a memory makes it easier to retrieve that memory again in the future. Critically for education, retrieving information during a test often leads to better future recall than rereading that same information, particularly if the final test occurs at some delay after the initial study session (Roediger & Karpicke, 2006a). The long-term benefits of testing make it a good candidate for use in education. However, testing sometimes does not help performance relative to restudying when a final test occurs shortly after studying, which may explain why students and teachers do not intuitively use testing as a learning strategy. That is, students learn that repeated studying (i.e. cramming) can get them through a test if the test occurs immediately after studying, but research shows that cramming leads to fast forgetting over time. Finally, testing does not just improve recall for factual information, but can also enhance the organization of information and the ability to transfer knowledge to new situations (for a description of ten benefits of testing, see Roediger *et al.*, 2011b).

As we shall show below, retrieval practice can be implemented in a variety of ways, both in the classroom and in students' own study routines. Testing also has indirect benefits for learning beyond directly enhancing retention of the tested information.

Types of Test

Students take many different kinds of tests and quizzes in the classroom, with a range of formats (including essay, short answer, and multiple choice) and levels of formality (including pop quizzes, weekly exams, mid-terms, and finals). The goal of most of these activities is to assess what students have learned from readings and lectures, but in some situations these assessments can also provide an opportunity for retrieval practice. Are some test formats (e.g. short answer or multiple choice) more effective than others at promoting learning?

On the one hand, McDermott *et al.* (2014) directly investigated the role of test format in a set of experiments that were conducted in real middle-school and high-school classrooms. In Experiment 4 the students took three short-answer or three multiple-choice quizzes about a particular topic covered in their history class. A research assistant administered the quizzes, so the teachers did not know what material was being covered in the quizzes. The day after the third quiz, the students took a unit exam consisting of a combination of multiple-choice and short-answer questions. Figure 6.1 shows that both multiple-choice and short-answer questions led to better performance on the unit exam compared with material that was not previously tested (other experiments in the study showed that the testing conditions consistently led to better performance on the unit exams than did restudying). Critically, the short-answer and multiple-choice test formats led to similar benefits in terms of performance on the unit exam. Furthermore, both quiz formats were equally effective regardless of whether the final test was short answer or multiple choice. Several other papers (e.g. Little *et al.*, 2012; Smith & Karpicke, 2014) have shown that multiple-choice and short-answer questions can lead to similarly large testing effects.

On the other hand, the two test formats are not equally effective under all conditions. Several studies have shown an advantage of short-answer questions over multiple-choice questions (Butler & Roediger, 2007; Kang *et al.*, 2007; McDaniel *et al.*, 2007). One explanation for this outcome is that short-answer questions, in which students are asked to generate the correct response, require more effortful retrieval than multiple-choice questions, in which students recognize and select the correct answer from among several alternatives. This is a prime example of what is known as a desirable difficulty in learning—some condition that is more difficult or effortful may seem to slow learning at first, but leads to memory benefits in the long term (Bjork, 1994; Karpicke & Roediger, 2007; Pyc & Rawson, 2009). Many researchers (e.g. Pyc & Rawson, 2009) have proposed that the difficulty involved in retrieval is the reason why testing is a powerful learning tool.

Retrieval effort is not the only factor that determines whether a question will lead to enhanced recall on a later test. Another factor is initial retrieval success. If students fail to answer a question on a practice quiz, they are unlikely to answer it correctly on a final test unless they get feedback or a chance to restudy. Not surprisingly, multiple-choice tests are

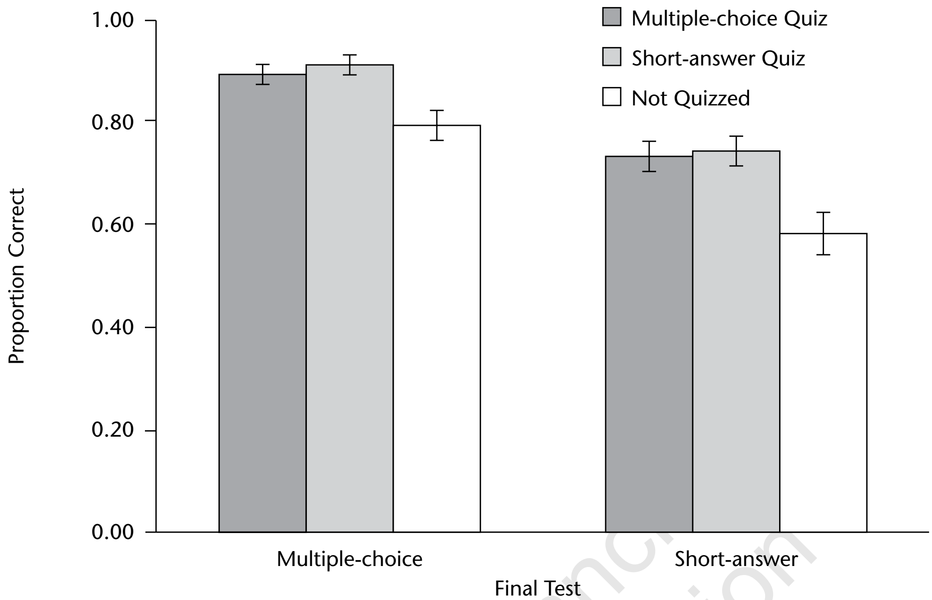

FIGURE 6.1 The results of Experiment 4 by McDermott *et al.* (2014). Both short-answer and multiple-choice quizzes led to enhanced performance on the final test (a unit exam) compared with a no-quiz control condition. The test format of the final test did not moderate the size of the testing effect.

Source: © 2013 American Psychological Association. Adapted with permission.

often easier than short-answer tests. Thus, in some experiments where performance on a short-answer practice quiz is low, multiple-choice questions can be more effective than short-answer questions (e.g. Kang *et al.*, 2007; Little *et al.*, 2012). Of course, one way to compensate for low initial test performance is to provide feedback after each question, which can lead to short-answer questions being more effective than multiple-choice questions (e.g. Kang *et al.*, 2007). We discuss feedback in more detail below.

One element that is important for multiple-choice questions is the construction of the questions themselves. Little *et al.* (2012) showed that properly constructed multiple-choice questions can enhance learning for both the correct and incorrect responses. Critically, the lures for the multiple-choice questions must be plausible and competitive; this requires students to retrieve information about why a particular response option is correct and why other response options are incorrect. In essence, a well-written multiple-choice question will require the test taker to retrieve information about all of the response options. Unfortunately, writing challenging multiple-choice questions is difficult, and many questions provided in test banks seem to have some implausible answers, which may not enhance future performance.

In summary, both multiple-choice and short-answer questions can be used to encourage retrieval-based learning, with much recent research suggesting that the two formats lead to similar testing effects. Both retrieval effort and retrieval success appear to be important in determining how effective a question is at engendering learning. Test questions should require some degree of retrieval effort, but if a question is too hard, performance feedback should be provided. On a related note, some research suggests that open-book tests (where students have access to their notes and other study materials) may be just as effective for promoting future learning as closed-book tests (Agarwal *et al.*, 2008).

Feedback

Although taking a test without feedback will enhance future recall, providing feedback (or providing students with the opportunity to restudy) almost always magnifies the size of a testing effect. Many aspects of feedback have been explored, such as how it affects later learning, what form the feedback should take (whether you simply say "right" or "wrong", or provide the correct answer), when feedback should be provided, and how taking a test allows people to learn more during their next study episode. We shall now consider each of these issues in turn.

As noted earlier, one important factor in finding a testing effect is initial test performance. If a student fails to answer a question correctly during the initial test, they are unlikely to answer that question correctly at a later date unless they are provided with feedback or given a chance to restudy (Butler & Roediger, 2008). Kang *et al.* (2007) had students read passages and then complete a short-answer or multiple-choice test before taking a final test later on. In one experiment they showed that when correct answer feedback was not provided, multiple-choice tests led to better final test performance than short-answer tests, because initial performance for the multiple-choice questions was better. When feedback was provided in a second experiment, however, that pattern was reversed, and short-answer tests led to better final recall than multiple-choice tests. Kang *et al.* attributed this reversal to the low initial test performance in the short-answer condition (compared with the multiple-choice condition) in the first experiment. Providing feedback (in the second experiment) allowed the students to correct any mistakes that they had made in the short-answer condition. Perhaps more interestingly, feedback can also help to strengthen correct responses, especially those made with low confidence on the initial test (Butler *et al.*, 2007, 2008).

Providing feedback is also important for multiple-choice tests, because when students take a multiple-choice test they are presented with a stem and several incorrect completions or answers along with one correct answer. The problem is that just reading (and definitely selecting) an incorrect lure can lead students to think that the response is correct, even if it is not. Roediger and Marsh (2005) showed that when students selected an erroneous response on a multiple-choice test and were later retested with a short-answer test, they often gave the incorrect response from the multiple-choice test, even though they were instructed to respond only if they were sure that their answer was correct. Furthermore, presenting additional lures on a multiple-choice test can decrease the size of a testing effect if performance is low (Butler *et al.*, 2006). Fortunately, providing feedback can ameliorate any negative effects of misleading lures, decreasing the likelihood that a wrongly endorsed lure will be reproduced on a later test (Butler & Roediger, 2008; see also Butler *et al.*, 2006; Marsh *et al.*, 2007).

Feedback can take many different shapes and forms (for a review, see Bangert-Drowns *et al.*, 1991). For example, verification feedback consists of telling the student whether their response was "right" or "wrong", whereas answer feedback provides the student with the correct answer after they have responded. Several studies have shown that, in general, simply telling students whether they are right or wrong leads to performance comparable to that observed when no feedback is provided (Fazio *et al.*, 2010; Pashler *et al.*, 2005; but for a discussion of how verification feedback can help with multiple-choice tests, see Marsh *et al.*, 2012). Thus, in general, feedback should at the very minimum involve providing students with the correct response.

Curiously, providing *more* information in a feedback message, such as explaining why an answer is correct, may or may not be helpful. Bangert-Drowns *et al.* (1991) conducted a meta-analysis and concluded that explanatory feedback did not provide any benefits over correct answer feedback. However, more recent research by Butler and colleagues has indicated that explanatory or elaborative feedback can be more effective than correct answer feedback *if* the final test includes inference or transfer questions (where students have to use knowledge in a new way). Butler *et al.* (2013) asked subjects to read a non-fiction text and then take a short-answer quiz in which they answered questions about the reading. After answering each question, students received either no feedback, correct answer feedback (where they were presented with the correct answer), or explanation feedback (where they received the correct answer along with more details about why that answer was correct). Two days later the students returned to take a final test in which half of the questions were repeated from the initial test and the other half required the students to make inferences based on knowledge that they had acquired from readings. The latter questions required students to extrapolate their knowledge. The results, shown in Figure 6.2, indicated that for the repeated questions the correct answer and explanation feedback conditions led to similar

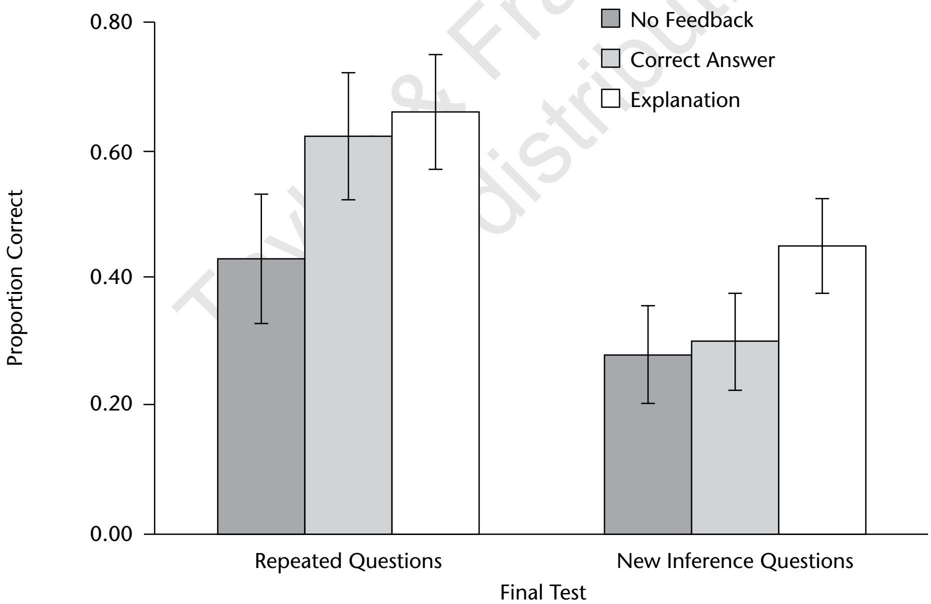

FIGURE 6.2 The results of Experiment 1 by Butler *et al.* (2013), Participants took an initial test with no feedback, correct-answer feedback, or explanation feedback. Performance on the final test (shown here) suggested that explanation feedback enhanced performance for inference questions compared with correct-answer feedback and the no-feedback condition. For the repeated questions, both correct-answer feedback and explanation feedback enhanced performance. Error bars are 95% confidence intervals estimated from Butler *et al.* (2013).

Source: © 2012 American Psychological Association. Adapted with permission.

recall on the final test, and that both led to better performance than the no feedback condition. In contrast, for the inference questions, only the explanatory feedback condition led to increased performance; recall on the final test was similar in the no feedback and correct answer conditions. Thus these results indicate that simple correct answer feedback is sufficient to improve performance when the final test questions are repeated from early tests, but that explanatory feedback can provide additional benefits to learning if the final test requires the making of new inferences.

When should feedback be provided? Early research on this question was murky (for a meta-analysis, see Kulik & Kulik, 1988). One view, grounded in behaviorist schools of thought, is that feedback should be provided as soon as possible after learning in order to reinforce correct responses and remediate incorrect ones. More recent research, however, has consistently shown that delaying feedback (by either a few seconds, minutes, or days) can be more effective than providing feedback immediately. For example, Mullet *et al.* (2014) conducted an experiment in a college engineering course in which students completed practice homework assignments throughout the semester. Sometimes students received feedback immediately after they had submitted their assignments, and sometimes they received feedback a week later. In two experiments, performance on the final course exam was better when students received delayed feedback than when they received immediate feedback. Mullet *et al.* suggested that the delayed feedback was more effective in promoting learning because it created a spacing effect. Another possible explanation was that the delay between the initial test and the presentation of feedback would allow the students to forget any incorrect responses, which would reduce interference with learning the correct response. Other research (Butler & Roediger, 2008; Metcalfe *et al.*, 2009) has provided further support for the view that delayed feedback is more effective than immediate feedback.

One caveat with regard to delayed feedback is that students must actually attend to the feedback in order to benefit from it. With delayed feedback there is the chance that students may simply look at their score on an assignment and not process the feedback related to any particular question. Mullet *et al.* (2014) required some students to look at feedback in order to receive credit for their assignments, and indeed students who were required to look at the feedback showed higher final test performance than those who were not required to do so.

Finally, one concept related to testing and feedback is test potentiation—the idea that people will learn more from reading a text if they have recently taken a test on material covered in the text, compared with people who have not recently taken such a test (e.g. Arnold & McDermott, 2013b; Izawa, 1966). That is, taking a test on material potentiates or increases future learning. Test-potentiated learning is different from the direct benefits of testing or of feedback—it is a benefit that accrues while reading something as a result of having recently taken a test. Arnold and McDermott (2013a) conducted an experiment in which students studied a set of 40 line drawings of objects before taking a final free recall test (writing down the names of the drawings). They were assigned to one of four conditions, with each group doing a different combination of practice tests and restudying before the final recall test. The experiment was a 2 × 2 between-subjects design: students took either no practice tests or three practice tests on the drawings, and they either restudied or did not restudy before the final test. All of the practice tests were free recall, and did not include feedback. The results (see Figure 6.3) showed a main effect of practice testing (more tests led to higher levels of recall on the final test), a main effect of restudying (restudying boosted recall on the final test), and, critically, a significant interaction. The interaction suggested that

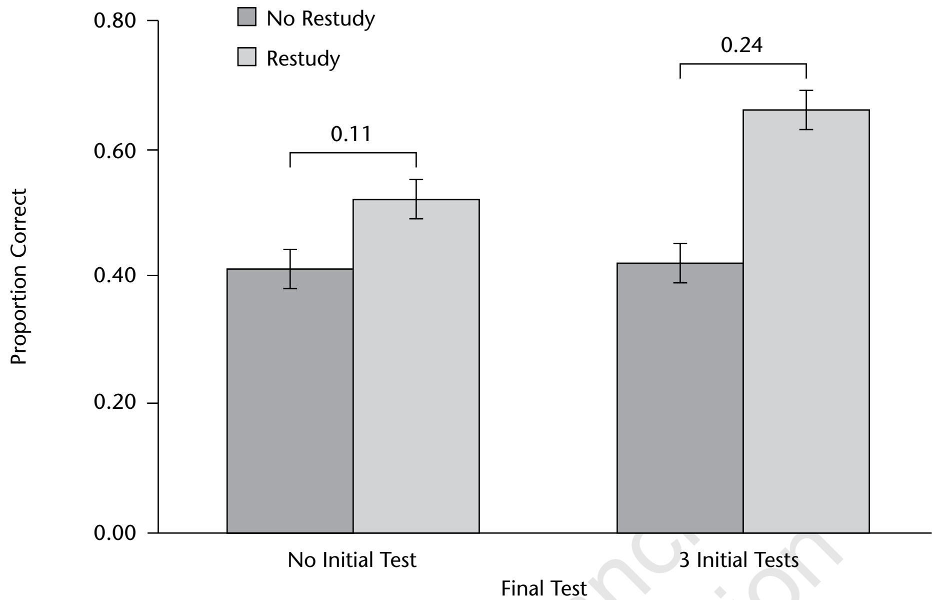

FIGURE 6.3 The results of Experiment 1 by Arnold and McDermott (2013a), showing the proportion recalled on the final test as a function of the number of previous tests and the restudy condition. Restudying after taking three tests led to a significantly better improvement than restudying after taking no initial tests. Error bars represent standard error of the means and are estimated from the original figure.

Source: © 2012 Psychonomic Society. Reprinted with permission.

students learned more from the restudy session after taking three practice tests than if they had not taken the practice tests. In other words, taking practice tests potentiated the students' ability to learn during the restudy phase.

Arnold and McDermott (2013a) suggested that test potentiation may occur because taking tests can lead to increased organization (Zaromb & Roediger, 2010), which helps people to learn more when they are restudying. However, taking a test can also have other positive effects on future studying. For example, Soderstrom and Bjork (2014) showed that students made better study decisions after taking a test, as they spent more time studying difficult word pairs compared with easy word pairs, and were more likely to study items that they had missed on the practice test. So testing helps to improve metacognition—students can learn what they know and what they don't know, and use that knowledge to guide their future studying.

To summarize, feedback is a valuable tool for use with retrieval-based learning. Feedback can help to correct mistakes and reinforce low-confidence accurate responses. At the very minimum, feedback should include the correct answer, but providing more detailed feedback can be helpful when the final test requires students to transfer knowledge. Finally, taking a test can help students to learn more from their next study session through test potentiation.

When and How to Test

By now we hope we have convinced you that tests (or retrieval practice) can be an effective way to improve learning. The next two sections in this chapter address some of the ways in which testing can be implemented, examining both the timing and dosage of tests (i.e. when to test and how many tests to administer), and how retrieval practice can be used both in formal tests and quizzes and in more informal ways. There are many different questions relating to how testing can be implemented, and we should also consider issues relating to differences among people (e.g. young children vs. adults) and topics (e.g. multiplication tables vs. history or chemistry). These considerations can lead to a dizzying number of combinations. Fortunately, however, research has revealed a few broad principles of test implementation that appear to be generally positive. We shall consider research that has documented several specific approaches to implementing tests in classroom environments.

The first principle is that more testing is better than less testing. Karpicke and Roediger (2008), for instance, reported an experiment in which students were asked to learn foreign-language word pairs. First the students learned the pairs relatively well. Then in some conditions students repeatedly retrieved pairs via testing, whereas in other conditions they repeatedly studied pairs. The results showed that after a 1-week delay, recall for the pairs that had been repeatedly retrieved was much better (80%) than recall for the pairs that had been repeatedly studied after being retrieved once (around 35%). Retrieving something on multiple occasions is much more effective than retrieving it once (see also Pyc & Rawson, 2009).

A second principle is that if an item or concept is going to be retrieved on multiple occasions, it is generally more effective if the retrieval practice is spaced apart in time, rather than being massed together (e.g. Pyc & Rawson, 2009). Massing tests will often lead to good performance on a test that occurs immediately after studying, but if there is a longer delay between studying and the final test, spacing practice tests apart will lead to better performance (another example of a desirable difficulty; Bjork, 1994). One extensively researched question concerns which schedules of spacing are most effective—that is, whether it is better for retrieval practice attempts to be spaced equally apart over time, or whether they should start immediately after study, and then occur less and less frequently over time (i.e. an expanding schedule). In general, most research suggests that expanding and equally spaced schedules are of similar effectiveness (Karpicke & Roediger, 2007; Logan & Balota, 2008; Kang *et al.*, 2014). In summary, more tests are better than fewer tests, and when multiple tests are administered, those tests should be spaced apart in time.

Memory researchers have also documented specific approaches to using retrieval practice in the classroom. One effective strategy is called the PUREMEM technique (Lyle & Crawford, 2011). This approach involved students taking a short quiz (5 to 10 minutes) at the end of every class, which covered content from that day's lecture (rather than the readings). Questions were displayed on PowerPoint slides and students wrote their answers on a sheet of paper (a clicker system or handing out paper quizzes would probably be equally effective). The course instructor reviewed the quiz at the start of the next day's class, and this provided an additional opportunity to learn the material. Overall, the quizzes accounted for 8% of each student's grade in the course, so the students took them seriously, but without having any individual quiz count for too much of their grade. The final tests consisted of four non-cumulative exams throughout the semester. As the PUREMEM quizzes targeted the most important content from the courses, questions from the quizzes did appear on the exams, but never as an exact repetition. Students in the PUREMEM classes scored an average of 86% in the exams, whereas students in control classes scored an average of 78% in the exams. Clearly, the quizzes were beneficial, although it is unclear whether the effect is a direct function of retrieval practice, or an indirect one (perhaps students studied more before each class and paid more attention when they knew that they would be quizzed every day). Other research has shown that practice quizzes can be effective even if they occur only weekly (McDaniel *et al.*, 2007).

Taking daily quizzes has been shown to have other benefits in addition to enhancing final test performance. Lyle and Crawford (2011) had students take a survey at the end of the course. Students in the PUREMEM class reported that the quizzing technique helped them to learn in a variety of ways—it allowed them to practice questions, encouraged them to come to class, and motivated them to pay attention in class. Thus the use of daily quizzes may have helped the students to realize the mnemonic benefits of testing.

Quizzing does not have to happen at the end of class. Leeming (2002) reported success with the use of quizzes at the start of class, whereas another approach is to include quizzes throughout a lecture. Szpunar *et al.* (2013) conducted an experiment in which students were asked to watch a 21-minute online lecture video about statistics. One group of participants took three 2-minute quizzes that were equally spaced throughout the lecture, whereas a second group of subjects did unrelated math problems instead of taking the quizzes. Performance on the cumulative final was better in the testing group than in the control group. In addition, students in the testing group took more notes, reported less mind wandering, and reported experiencing less anxiety when going into the final exam. The findings of a subsequent study (Szpunar *et al.*, 2014) suggested that taking the interpolated quizzes led to better metacognitive monitoring—students were more accurate in their predictions of their future test performance after taking a test. Thus the inclusions of short quizzes during a lecture can improve later test performance, encourage students to pay attention in class, and improve metacognition. The optimum spacing of such quizzes still needs to be clarified. Schacter and Szpunar (2015) have provided a teacher-friendly review of many of these issues.

Finally, we shall end this section by encouraging the use of cumulative tests. Although students typically dislike cumulative exams, testing material from the entire semester provides an opportunity for spaced retrieval practice. Carpenter *et al.* (2012) have recommended that in addition to final exams being cumulative, unit tests should be cumulative throughout the semester. Not only does this provide spaced retrieval practice, but also students who know that they will be taking a cumulative exam will study material from the entire course before each test. Using such exams will make it more likely that students will remember course content beyond the current semester. The aim of education is (or should be) for long-term learning to occur, rather than just "getting through the test on Friday."

Formal and Informal Testing

Retrieval practice can work with any activity that requires students to retrieve information from memory. Although formal tests and quizzes in the classroom have provided one of the most straightforward ways to implement retrieval-based learning, retrieval practice can also be used in more informal ways. In both the PUREMEM approach described above (Lyle & Crawford, 2011) and the experiments conducted by Szpunar *et al.* (2014), the main goal of the quizzes is not to evaluate students, but rather to increase classroom engagement and encourage learning. Retrieval practice can also be effective when it is integrated into classroom assignments and discussion.

One recent study (Lindsey *et al.*, 2014) had eighth-grade Spanish students use a computerized review system during class time. The computer program cued students with an English word or phrase and asked them to recall the Spanish translation. Whether or not they were correct, students saw feedback. The students used this computerized flashcard system each week to study material from 10 different chapters over the semester. Some of the material was studied in a massed fashion (in which students only studied material from the current week), some of the material was studied in a generic spaced fashion (in which students studied a combination of material from the current and previous weeks), and finally some of the material was presented in a personalized spaced fashion (a computer algorithm selected specific items for students to study based on the item difficulty, the student's ability, and how often the student had seen the item before). In a cumulative final exam, the material that was studied with the personalized review led to the best performance (73%), followed by the generic spaced schedule (67%), and the worst performance (although not by much) was for the massed study condition (65%). As the authors noted, this finding is striking, in view of the fact that the manipulation only required 30 minutes of time per week (about 10% of the time for which students were engaged with the material overall), and students were free to spend as much time as they wished studying, paying attention in class, and doing additional reading. Despite the lack of a non-tested control condition, these results suggest that retrieval practice can be effective when used as a classroom activity, rather than a formal quiz.

Another simple way to use informal retrieval practice in the classroom is to ask students questions. Obviously most teachers present questions to their class every day, but with a little prior thought they can ask questions in a way that promotes retrieval practice for all of their students. Previous research suggests that people can derive benefit from retrieval practice even if they only *think* about the answer to a question—writing the answer down or saying it aloud is not necessary (Putnam & Roediger, 2013; Smith *et al.*, 2013). So how can a teacher ask questions in a way that encourages students to covertly answer them? One approach promoted by Pashler (personal communication, 24 September 2013) is called the "On the Hook" procedure. In this approach, each student in the class has their name written on a Popsicle stick, which is kept in a coffee can. The teacher asks the class a question and allows a few moments for everyone to think of an answer. Then, instead of calling on a student who has their hand raised, the teacher pulls a stick at random from the coffee can and asks that student to answer the question. Thus every student has a chance to covertly answer the question, and is motivated to do so because they know that they might be called upon to answer the question aloud in front of the class. Pashler reported that this technique is effective, and in at least one case it has been adopted by an entire school with huge success.

Retrieval Practice as a Study Strategy

Students can also use retrieval practice when they are studying. Unfortunately, survey studies suggest that students of many different ages report using and preferring relatively ineffective study strategies, such as highlighting or rereading, rather than using effective strategies such as retrieval practice and spaced study (Agarwal *et al.*, 2014; Hartwig & Dunlosky, 2012;

Karpicke *et al.*, 2009; Kornell & Bjork, 2007). Here we highlight some of the ways in which students can use retrieval practice in their own study routines.

First, students may already use retrieval practice with flash cards (Kornell & Bjork, 2008). When these are used correctly, students should look at the cue on one side of the card and attempt to recall the answer before looking at the other side of the card for feedback. In other words, they are practicing covert retrieval. One important idea to note here is that most students stop studying too soon—they remove a card from the stack once they have recalled it correctly once (Karpicke, 2009; see also Kornell & Bjork, 2007, 2008). As discussed throughout this chapter, students often have a poor understanding of their own learning. Therefore they should be encouraged to successfully retrieve information more than they think is necessary. Rawson and Dunlosky (2012) recommend recalling an item three times during initial learning, and then learning it again during three future study sessions.

Second, with some foresight students can turn other activities into opportunities for retrieval practice. For example, students are sometimes asked to create concept maps, where they use circles and lines to portray diagrammatically the relationships between different ideas (Novak, 1990). The creation of concept maps is often touted as an effective study strategy, and it does force students to think about the meaning of material and the relationships between concepts. However, Karpicke and Blunt (2011) showed that creation of concept maps was less effective than straightforward retrieval practice, where students are simply asked to recall everything they can remember from a chapter. Fortunately, a simple shift in procedure can turn concept mapping into a relatively effective study aid. Rather than creating a concept map while looking at a textbook, students should attempt to create concept maps from memory, and then consult the textbook after they have finished a first draft of the map. In this way, creating the concept map becomes a form of retrieval practice, and rereading the text after doing so may lead to test-potentiated learning. Blunt and Karpicke (2014) conducted an experiment in which they had students read a text and then recall information either by writing a paragraph about what they had learned (the standard approach to retrieval practice), or by creating a concept map from memory. No feedback was provided in either condition. The results showed that both conditions led to similar positive effects on a delayed test.

Finally, students can use various forms of retrieval practice to improve their recall of text materials. In one system, called the Read–Recite–Review strategy (or 3R strategy), students are asked first to read the chapter, then to take a few minutes to recall aloud everything they can remember, and then to re-skim the chapter to evaluate how well they did. This strategy enhances learning as assessed on both immediate and delayed free recall tests, compared with simply rereading or taking notes, and with some types of materials it can enhance performance on multiple-choice tests as well (McDaniel *et al.*, 2009). Given that rereading is a favorite study strategy of students (Karpicke *et al.*, 2009), the 3R strategy might be an important and useful technique for them to know.

In summary, students do not have to rely on their teachers to formally quiz or test them in order to derive benefits from retrieval practice, as they can also use such techniques in their own study sessions.

Indirect Benefits

So far the discussion has primarily focused on the direct memorial benefits of testing—retrieving something from memory aids learning. However, frequent testing or quizzing can have additional side effects. We have already considered a few examples, such as how interspersing quizzes throughout a lecture can enhance attention and reduce mind wandering (Szpunar *et al.*, 2014), or how cumulative exams can encourage students to study material from the entire semester before each test (Carpenter *et al.*, 2012). Roediger *et al.* (2011b) explored ten different benefits of testing (see Table 6.1). The first benefit was the direct benefit, and the other nine benefits were positive side effects that can occur after using frequent, low-stakes quizzes in the classroom, such as the fact that testing can help students to transfer knowledge to new situations. Here we highlight two of the benefits covered by Roediger *et al.* (2011b), and describe an additional indirect benefit.

One benefit, already noted earlier, is that testing improves metacognition and allows students to identify gaps in their own knowledge. As discussed in the introduction, students have a tendency to reread material rather than test themselves, both because rereading seems easier than quizzing oneself, and because rereading can lead to increased feelings of fluency, or knowing (Karpicke *et al.*, 2009). This can be dangerous, because students will often report being more confident about having learned something after repeatedly reading, even though such confidence is unwarranted (e.g. Roediger & Karpicke, 2006a). For this reason, practice tests are important—students will realize what they know and what they do not know, and will update their predictions of their performance appropriately. For example, Szpunar *et al.* (2014) showed that taking one quiz lowered students' expectations of their future performance to match their actual future performance, whereas taking three tests raised their performance on future tests to match their initially overly confident estimates. Other research has shown that, after taking a quiz, students spend more time studying material that they initially answered incorrectly (Son & Kornell, 2008), and also that they spend more time studying more difficult material (Soderstrom & Bjork, 2014). Although students tend not to use self-testing while studying, Kornell and Bjork (2007) reported a survey which suggested that when students did test themselves, 68% of the time it was in order to measure what they had learned (they did not seem to know about the direct benefit of retrieval practice). Thus using frequent tests and quizzes in the classroom can help students to identify what material they know and what they do not know, as well as help them to make more accurate predictions about the state of their own learning.

TABLE 6.1 The ten benefits of testing listed by Roediger *et al.* (2011b). Benefit 1 refers to the direct benefit of testing, whereas Benefits 2–10 refer to positive indirect benefits of testing.

Benefit 1	The testing effect: retrieval aids later retention.
Benefit 2	Testing identifies gaps in knowledge.
Benefit 3	Testing causes students to learn more from the next learning episode.
Benefit 4	Testing produces better organization of knowledge.
Benefit 5	Testing improves transfer of knowledge to new contexts.
Benefit 6	Testing can facilitate retrieval of information that was not tested.
Benefit 7	Testing improves metacognitive monitoring.
Benefit 8	Testing prevents interference from previous material when learning new material.
Benefit 9	Testing provides feedback to instructors.
Benefit 10	Frequent testing encourages students to study.

Source: Roediger *et al.* (2011b). Reprinted with permission.

A second important benefit of frequent classroom quizzing is that it encourages students to study more often. As you might know from your own experience and that of friends, most students report that they do the majority of their studying the night or day before an exam (Mawhinney *et al.*, 1971). Agarwal *et al.* (2014) surveyed middle-school and high-school students about their study habits in different classes. Critically, in some of the classes students were participating in experiments where retrieval practice was integrated within the course (e.g. McDermott *et al.*, 2014). The results of the survey showed that across all grade levels and topics, students studied for approximately 19 minutes per week outside of class when there was no test scheduled for that week, but they studied for 43 minutes per week outside of class when a test was scheduled. Thus, not surprisingly, integrating daily or weekly quizzes into a class can encourage students to study more often and more consistently throughout the semester.

One additional indirect benefit of testing which was not listed by Roediger *et al.* (2011b) is that frequent classroom testing can reduce test anxiety. Agarwal *et al.* (2014) had students complete a survey at the end of the year, in which they answered several questions about testing and taking clicker quizzes in class. All of the students were included in at least one classroom study that used retrieval practice (e.g. McDermott *et al.*, 2014). The most salient finding was that 72% of the 1,408 middle-school and high-school students reported that taking the practice tests made them less anxious about unit exams, whereas 22% said that they were equally anxious, and only 6% said that they were more anxious. Thus, in contrast to the assumption that quizzing students more often may increase test anxiety, these results suggested that frequent quizzing actually reduces test anxiety, either by giving students more practice in taking tests, or by helping them to learn the material better.

In summary, including frequent low-stakes quizzes can have a variety of benefits in the classroom. In addition to the direct memorial benefits of retrieval practice, frequent quizzes can enhance students' metacognition of their own learning, encourage them to study more, and decrease test anxiety.

Distributed Practice and Spacing

One lesson is typically insufficient to create learning that lasts over time. Therefore repetition is a necessity in education, and so the issue of *when* students should restudy is central to instruction. When students study material more than once (or when a teacher reviews material), the timing of the subsequent learning sessions is important. When a student reviews a critical section of a textbook, should it be soon after the first reading, or should she wait a week? When a teacher plans review sessions to prepare his students for the final exam, should the review of a topic be included at the end of the lesson for that topic, or should he spend time today reviewing material covered a month ago? A century of research (Ebbinghaus, 1885/1964) suggests that students should space their repetitions of learning over time, and that longer spacing gaps are more effective than shorter ones. Distributing practice over time enhances learning. Below we review research on distributed practice, focusing primarily on research conducted with educationally relevant materials, learners, settings, and time scales.

What is the Spacing Effect?

Repeated sessions of study spaced over time lead to more effective learning than repetitions that occur back to back. This finding is referred to as the spacing effect or the distributed practice effect (Cepeda *et al.*, 2006). The term lag, or spacing gap, refers to the amount of time (or sometimes the number of other episodes) that elapses between two episodes of learning a piece of information. Repetitions with a lag of zero constitute massed practice (back to back), whereas distributed (or spaced) practice includes any lag greater than zero. For example, a student learning vocabulary might choose to repeat a word-definition pair in massed fashion (e.g. "paucity–shortage", "paucity–shortage", "paucity–shortage"). Alternatively, he could use a spaced practice schedule, by studying this pair with other vocabulary pairs between repetitions (e.g. "paucity–shortage", "loggia–balcony", "sobriquet–nickname", "paucity–shortage"). This constitutes within-session spacing (e.g. Dempster, 1987). Another option would be to study the vocabulary pair(s) across many days (e.g. once a day on Monday, Wednesday, and Friday). This constitutes between-session spacing (Küpper-Tetzel, 2014). In this example, both types of spacing (within-session and between-session) would lead to better long-term learning than massed practice. Researchers sometimes distinguish the spacing effect, which suggests that distributed practice is more effective than massed practice, from the lag effect, which suggests that longer lags between repetitions promote more durable learning than shorter lags (Crowder, 1976). However, for the purposes of this review we shall use the terms "distributed practice" and "spacing" to describe all variations of these basic findings.

The simplest design for evaluating distributed learning is shown in Figure 6.4, and includes (1) initial learning, (2) relearning, (3) a lag between the two learning sessions, (4) a final test, and (5) a retention interval between relearning and the final test. Note that the overall amount of time spent studying is equated across conditions, so any differences in performance on the final test must be due to the *distribution of time* spent studying, rather than to the *amount of time* spent studying. The key manipulation in this type of experiment is that of the spacing gap (i.e. the lag between the two learning trials), but as we shall see later, the length of the final retention interval is also important.

As noted earlier, spacing can be manipulated within sessions or between sessions (Küpper-Tetzel, 2014). Within-session spacing occurs when the lag between items in a series is manipulated by inserting other items between repetitions. Thus, in the within-session paradigm, spacing occurs on timescales ranging from seconds to minutes. One specific method that induces within-session spacing is called interleaving (Rohrer, 2009), which consists of mixing up practice with related types of materials. Practice problems in mathemematics textbooks are often *not* interleaved—students typically practice a block of problems about one topic (e.g. calculating the volume of a spheroid), and then in another

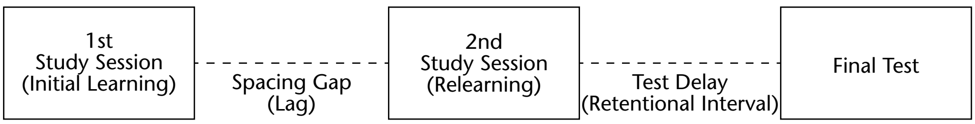

FIGURE 6.4 Design of a typical spacing experiment. Participants study information in two sessions that are separated by an interval ranging from zero (massed practice) up to years (but typically shorter than that). This interval is referred to here as the spacing gap (or lag). Participants are finally tested after what is referred to as a test delay (or retention interval). The spacing gap is typically manipulated, and the retention interval is manipulated in some (but not all) studies.

section practice a block of problems about a different topic (e.g. calculating the volume of a spherical cone). In an interleaved schedule, different kinds of problems would be mixed together (e.g. students might calculate the volume of a spheroid, then a half cone, then a wedge). Although blocked practice is common in education, it may be less effective than interleaving. Rohrer and Taylor (2007) demonstrated this by showing that interleaving practice problems led to better performance on a final mathematics test (63%) than blocked practice (20%), even when the test occurred after a 1-week delay. Despite the fact that students often report feeling that massed practice is more effective than interleaving, the benefits of interleaving have been shown in many domains (see Rohrer, 2009; see also Chapter 5 in this volume).

In contrast to within-session spacing, between-session spacing occurs when material is studied in one session and then covered again in a second session. The lag between the sessions is filled with something completely unrelated to the target material (e.g. taking a nap, working on a different class, or simply doing nothing), and can be as short as a few minutes between lists or as long as a few years (Bahrick *et al.*, 1993). Both within-session and between-session spacing can enhance learning compared with massed practice.

The Benefits of Distributed Practice

Although the advantage of distributed practice over massed practice has been demonstrated most commonly on retention tests involving college-aged students learning discrete verbal materials, many studies have extended the effect beyond these basic conditions. The spacing effect has been shown to occur in many different animal species and at nearly every stage of human development. It has also been demonstrated with a wide range of learning materials and across multiple measures of learning (for an in-depth review of the breadth of the spacing effect, see Gerbier and Toppino, 2015). Although researchers are still debating why exactly the spacing effect occurs (for reviews, see Delaney *et al.*, 2010; Hintzman, 1974), the robustness of the effect makes it a prime candidate for educational applications.

One notable boundary condition for the spacing effect is that massed study can sometimes be more effective than spacing when the final test occurs immediately after the last study session. Balota *et al.* (1989) reported an experiment in which subjects took a test either immediately after a second presentation of a paired associate or after a delay. With the immediate test, massing led to better recall than spacing, but with the delayed test, spacing led to better recall than massing. Although this finding is an important boundary condition of the spacing effect, in educational scenarios students are rarely tested immediately after study.

Spacing Makes It Stick

Having reviewed the basic research on the spacing effect, we shall now examine research that has used educationally relevant materials and studies conducted in real classrooms.

Translational Laboratory Research

Many investigators have recently begun exploring spacing effects with educationally realistic materials. Fortunately, the results generally corroborate basic research in showing advantages for distributed learning. There are many studies like this, but here we highlight two which show that spacing can enhance multiple forms of learning.

The first, by Gluckman *et al.* (2014), is a laboratory study of elementary school students that examined whether spaced practice would enhance both their memory and their ability to generalize science concepts to new domains. The experimenters presented first- and second-grade children with four lessons about food chains. Each lesson covered a different biome (grasslands, arctic, ocean, and swamp), but key concepts were repeated across lessons (e.g. the definition of a predator). The children either completed all four lessons in a single day (the massed condition) or received one lesson on each of four consecutive days (the spaced condition). One week after the last lesson, the students took a final test that assessed memory for facts (e.g. what a carnivore eats), simple generalization (e.g. larger animals typically eat smaller animals), and complex generalization (e.g. animals in a food chain are dependent on one another for food and survival). Critically, the memory questions corresponded directly to facts that the students learned in the lessons, whereas the generalizations (both simple and complex) were taught with one set of animals during the lessons, but tested with a new set of animals from a novel biome (e.g. desert). Figure 6.5 shows performance on the final test, and reveals that the spaced condition led to betterperformance on all three kinds of test than did the massed condition. Thus spacing enhanced both memory and transfer.

In another study using realistic materials, Kapler *et al.* (2015) examined the effects of spaced practice on memory and higher-order learning. They simulated a science class by having introductory psychology students watch a lecture about meteorology in a lecture hall. This design provided more external validity than is typically present in a laboratory, but allowed the researchers to control other variables, such as studying outside of class (the meteorology lecture was excluded from students' grades). After the initial lecture, students completed an online review module 1 day or 8 days later, and then took a final test in class 35 days after the review module. The online review session incorporated testing to capitalize on the benefits of retrieval practice. Critically, the review module and the final test contained both factual questions (in which students simply had to recall a fact) and higher-order questions (in which students had to apply a concept to a new problem). The 8-day spacing gap led to better final test performance than the 1-day spacing gap for both factual questions (54% vs. 47%) and higher-level questions (43% vs. 36%). Thus, in an experiment involving fairly realistic materials, delays, and tests, spacing led to an increase in performance of 7%, or the equivalent of half a letter grade.

Research in Instructional Settings

Spacing effects have also been shown to enhance memory and other forms of learning in real classrooms and in other training environments. For example, Sobel *et al.* (2011) had students in a fifth-grade classroom learn GRE vocabulary words and definitions (e.g. "gregarious: outgoing and social") in two learning sessions that were spaced by either 1 minute (massed condition) or 7 days (spaced condition). Five weeks after the second learning session, students were given the vocabulary words and asked to recall the definitions. Not surprisingly, recall was better after spaced practice (20.8%) than after massed practice (7.5%), demonstrating that middle-school students can benefit from distributed practice in their normal classroom environments (see also Carpenter *et al.*, 2009).

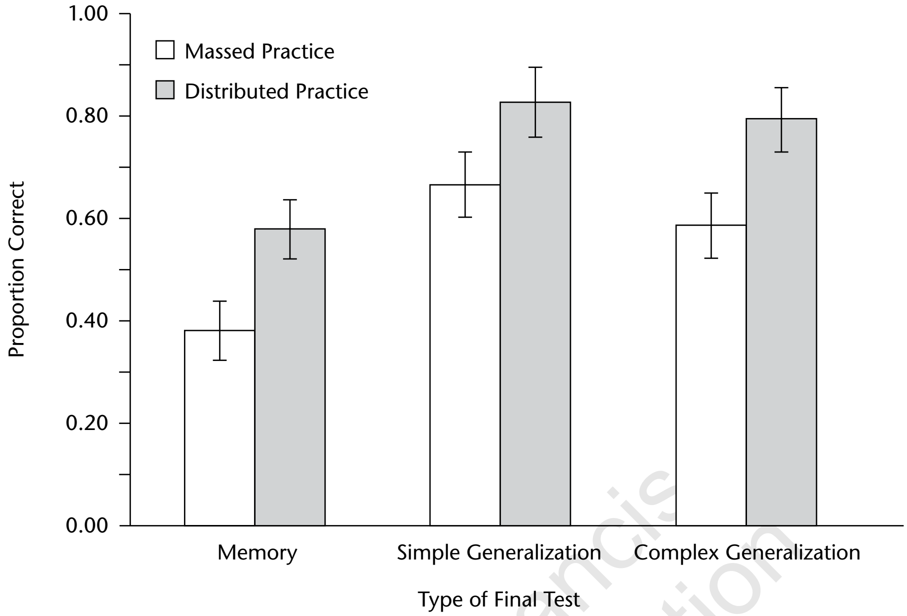

FIGURE 6.5 Performance on the final test in the study by Gluckman *et al.* (2014), expressed as proportion correct on three types of test (memory, simple generalization, and complex generalization), plotted as a function of practice condition (massed practice vs. distributed practice). In this experiment, early elementary school children were taught about food chains in four lessons that were either distributed across 4 days (distributed practice) or took place over the course of 1 day (massed practice). They were then tested 1 week after the last lesson. Distributed practice led to significantly better performance on all three test types. Error bars are standard error of the mean (calculated from Table 1 in the original report; Gluckman *et al.*, 2014, p. 270).

Source: © 2014 John Wiley & Sons Ltd. Adapted with permission.

Moving beyond basic learning, Bird (2011) manipulated the spacing of practice sessions in which students learned English syntax in a college-level second-language learning course for native Malay speakers. Students practiced identifying sentences with syntax errors (e.g. "I have seen that movie with my brother last week"), and were given feedback on their performance in practice sessions that occurred either 3 or 14 days apart. The final test, which was 7 or 60 days later, required students to read novel sentences that were not part of the learning phase and mark whether the sentence used correct syntax. Thus the test required the abstraction of syntactical rules rather than rote memory. When the retention interval was 7 days, the 3-day and 14-day conditions led to similar performance on the final test, but when the retention interval was 60 days, the 14-day condition led to enhanced performance compared with the 3-day condition. Bird's study demonstrated that spaced practice can promote transfer of learning to new material, and that this benefit persists over time.

Finally, researchers have also examined the benefits of distributed practice in training domains outside standard classrooms. Efficient training that leads to durable learning is highly valued in many fields, including medicine and police work, particularly because the skills required in such fields require advanced forms of learning. For example, the Enhanced

Cognitive Interview (ECI) is a police interviewing technique originally crafted by cognitive psychologists to standardize eyewitness interrogation procedures while reducing the influence of interviewer bias and other related pitfalls (Fisher & Geiselman, 1992). The ECI has been adopted by many police organizations, yet there have been some problems with regard to questions of how best to train new officers in the ECI approach. A recent study by Heidt *et al.* (2016) revealed that distributed practice holds promise for teaching the ECI. In their experiment, 60 participants were given 2 hours of training on the ECI, either in a single 2-hour session (massed practice) or in two 1-hour sessions with a spacing gap of 1 week between sessions (spaced practice). The participants in the spaced practice condition showed much greater improvement in multiple aspects of using the ECI technique. For instance, spaced practice led to greater use of open-ended (non-suggestive) questions, which is a core principle of the ECI. In short, the simple change of distributing practice—rather than massing—has the potential to change the way in which eyewitnesses are interviewed by police, which in turn could affect the quality of evidence used in the pursuit and prosecution of criminals. Moulton *et al.* (2006) provided a similar example of how spacing practice sessions for medical students enhanced their ability to learn a difficult and dangerous microsurgical technique. In short, distributed practice enhanced a training program for a complex motor skill that can save lives.

The Optimal Spacing Gap

So far, we have illustrated that spacing typically outperforms massing. But how long should the spacing gap be? Is 5 minutes enough or should learning sessions be spread out as far as possible? A complete answer undoubtedly depends on a variety of factors (e.g. what is being studied, who is studying, how learning will be assessed, etc.). One factor that does appear to be important is how long learning needs to be maintained, or the length of the retention interval between the last study session and the final test. As we noted earlier, one boundary condition of the spacing effect is when the retention interval is very short. Indeed, recent research suggests that an optimal spacing lag depends on the length of the retention interval, and that longer retention intervals require longer spacing gaps (although, as with any rule, there are exceptions).

In what is almost certainly the most comprehensive examination to date of the effects of spacing gaps as assessed after various retention intervals, Cepeda *et al.* (2008) recruited 1,350 people online for a long-term investigation of spacing effects (for a similar laboratory study, see Cepeda *et al.*, 2009). In this experiment, the participants studied a set of obscure but true trivia facts (e.g. "What European nation consumes the most spicy Mexican food?" Answer: "Norway"). The participants then reviewed the trivia facts after lag periods of 0, 1, 2, 4, 7, 11, 14, 21, or 105 days following the initial learning (different groups were given different lag periods). Finally, the participants were tested at retention intervals of 7, 35, 70, or 350 days after the review session. The results are somewhat complex, but two conclusions seem unequivocal. First, non-zero spacing gaps produced better learning than did massed practice—across retention intervals, the optimal gap improved recall performance over the massed condition by 64%. Second, and more relevant for the current discussion, there was a different optimal gap at each retention interval. For example, the optimal spacing gap at the 350-day retention interval was 21 days, whereas at the 7-day retention interval it was only 1 day. Thus the simple conclusion that longer spacing gaps between presentations are always better for performance is wrong. Instead, the optimal spacing gap depends on the length of the retention interval.

Critically, this pattern has also been shown with educationally relevant materials. Rawson and Kintsch (2005) had students read a 1,730-word text from a *Scientific American* article and then reread the text either immediately or after 1 week. The final test took place either 5 minutes after or 2 days after the second learning session, and consisted of both a free recall section and a series of short-answer questions that required the making of inferences and application. Figure 6.6 shows the students' performance on the short-answer questions. At the immediate test (5-minute retention interval) the massed practice condition led to similar performance to that produced by the spacing condition, but at the delayed test (2-day retention interval), spaced practice led to better performance than massed practice. The same pattern occurred in the free recall section of the test. Similar results were reported by Verkoeijen *et al.* (2008), who showed that when the retention interval between the second study session and the final test was 2 days, a spacing gap of 4 days led to better performance than either massed rereading or a spacing gap of 3.5 weeks. Taken together, these studies show first that massed rereading appears to be advantageous only when the test is administered immediately after studying, and second, that spacing gaps that are too long can sometimes be ineffective.

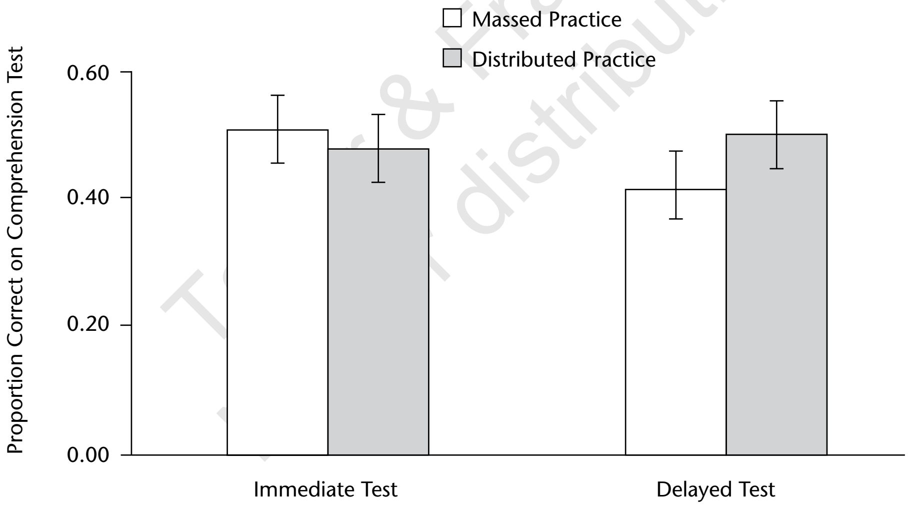

FIGURE 6.6 The results of Experiment 2 by Rawson and Kintsch (2005), showing performance as a proportion of correct scores on a comprehension test consisting of short-answer questions, plotted as a function of practice condition (massed practice vs. distributed practice) and test delay (immediate test vs. delayed test). In this experiment, the participants read a passage of text twice, with the readings separated by a lag of either zero or 1 week, and they were then tested either 5 minutes or 2 days after the second reading. Performance was not significantly different between the practice groups on the immediate test (although it slightly favored massed practice), but spaced practice produced higher comprehension scores on the delayed test. Error bars are standard error of the mean (K. Rawson, personal communication, 25 August 2015).

Source: © 2005 American Psychological Association. Adapted with permission.

Finally, the finding that retention interval influences the optimal spacing gap between presentations has also been demonstrated in a classroom field study (Küpper-Tetzel *et al.*, 2014). German sixth-grade students studied German–English vocabulary in an initial session, and they then studied the same vocabulary terms in a second session either immediately, 1 day later, or 10 days later. The final test took place 7 or 35 days after the second session. In line with the laboratory studies reviewed here, the optimal spacing lag at the short (7-day) retention interval was 1 day, whereas the 1- and 10-day spacing gaps were equally advantageous over the massed practice condition when the retention interval was longer (35 days).

There has now been enough research—in the laboratory and in the classroom—to allow some initial conclusions to be drawn about the relationship between spacing gaps and the retention interval. Clearly, students should avoid massed repetitions, as these appear to be of little or no benefit to long-term retention (in fact, they are sometimes no better than a single exposure to material; e.g. Callender & McDaniel, 2009; Rawson & Kintsch, 2005). Once spacing is introduced, though, a rough guideline is that longer retention intervals generally require longer spacing gaps. However, there is a limit—for any given retention interval, test performance is an inverted U-shaped function such that increasing the spacing gap first increases performance until an optimal lag, after which performance declines with increasing lag periods. Determining an optimal spacing gap for any given setting would require additional research, but the available data suggest that the optimal gap is often 5–40% of the retention interval, and that this ratio gets smaller as the retention interval gets longer. For example, Cepeda *et al.* (2009) found that when the retention interval was 1 week the optimal gap was 14% (1 day), whereas when the retention interval was 1 year the optimal gap was 6% (21 days).

Unfortunately, this makes planning for practical purposes more difficult. Teachers cannot simply plan a relearning session far in the future in the hope that the longest spacing gap will lead to the best possible performance. To optimize performance, the retention interval must be taken into consideration. One way to conceptualize this is that, after a very long delay, sufficient forgetting has occurred for rereading to be functionally equivalent to reading the material for the first time. However, educators and students should be able to work with these estimates to roughly determine what is a good spacing schedule. One final caveat is that it might be better to err on the side of a spacing gap that is too long rather than too short, as the cost (in terms of lowered test performance) of an overlong spacing gap is much smaller than the alternative (Rohrer, 2015).

How to Use Spacing

In summary, the timing of learning sessions can have powerful effects on retention even when the overall time on task is equated. The spacing of repetitions nearly always enhances long-term learning, and longer spacing gaps (but not too long!) typically lead to more durable learning. Empirical research suggests that these effects apply to numerous types of learning, in many environments, with many kinds of students. Here are a few recommendations for using distributed practice in the classroom.

First, instructors can add review sessions at the start of each lesson where they cover key concepts from previous lessons. In many cases, old content can be used to introduce new concepts, perhaps by comparing and contrasting related topics, or by pointing out connections between seemingly unrelated concepts. Class time is obviously limited, so these reviews should be brief and focus on the most important concepts from a lesson. A second strategy that serves a similar purpose without taking up class time is the setting of homework assignments that target both old and new material. In this way students can both practice new concepts and refresh their understanding of old concepts. Third, instructors should keep retention interval in mind when considering when to schedule review sessions. Longer retention intervals require longer spacing gaps (but be careful of overlong gaps). In schooling situations, we suspect that spacing gaps of weeks or months are ideal. Finally, one particularly effective strategy is to pair distributed practice with retrieval practice. As noted earlier, one good way to do this is to administer cumulative (or semi-cumulative) exams. For example, in a psychology research methods course (taught by the second author), each of the five exams conducted during the semester consisted of a mixture of old (20%) and new (80%) content. The final exam covered the entire course. With semi-cumulative exams students are given spaced exposure, get retrieval practice, and are motivated to study lessons from the entire course.

Conclusion

In summary, both retrieval practice and distributed practice are powerful learning tools that can enhance performance in a variety of situations. In reviewing the literature we have attempted to summarize basic research, research with educationally relevant materials, and research conducted in real classrooms. Fortunately, the basic principles discovered in the laboratory appear to translate readily to the classroom. In an effort to continue the translational research cycle, we shall close with a few practical applications of retrieval practice and spacing. In addition to being effective strategies, these suggestions are easy to implement, and require little or no extra equipment or resources.

From the Laboratory to the Classroom

Retrieval Practice

- One of the easiest ways to integrate retrieval practice into a study routine is by using flashcards. Flashcards can be used to remember vocabulary words, definitions, and more. To use flashcards correctly, it is important to attempt to retrieve the answer before turning the card over. In addition, spacing effect research (e.g. Kornell, 2009) suggests that using one larger stack of flashcards is more effective than using several smaller stacks.
- Teach students to use the Read–Recite–Review method (McDaniel *et al.*, 2009) when reading textbooks and other written materials. Briefly, students should read a chapter and then spend a few minutes recalling aloud everything they can remember from the passage (they could also write down their responses). The students should then review the chapter to see what they missed and what they remembered.
- Low-stakes in-class quizzes are an effective way to boost grades in a course (Lyle & Crawford, 2011). Quizzes can be given at the beginning or end of class, and can be administered with a clicker system or using pencil and paper. The quizzes should not represent a large proportion of a student's grade, but should be worth a few points so that

students take them seriously. Daily quizzes can directly improve memory for the tested material, and will encourage students to keep up with the reading and to ask questions in class. Reviewing the quizzes at the start of the next class session builds in a spaced presentation of the material.

- Instructors can use the "on the hook" method to encourage covert retrieval practice in their classrooms. Each student's name is written on a Popsicle stick and put in a coffee can. When the teacher asks a question she poses it to the whole class, waits a few seconds, and then draws a name from the coffee can and asks that student to answer the question. This structure means that most students will covertly retrieve an answer so that they can respond if called upon to do so.
- Finally, instructors can take advantage of technology to integrate short quizzes into material presented online. Short quizzes inserted in a lecture video posted online can reduce mind wandering and reduce the amount of anxiety that students feel about a final exam (Szpunar *et al.*, 2013). Alternatively, instructors could post short online quizzes (perhaps via a course management system such as Blackboard or Moodle) that tests material from both readings and class lectures.

Distributed (Spaced) Practice

- Instructors can provide brief reviews of previously covered content at the start of each new lecture. Although it may be somewhat useful to briefly remind students of the most recent material (e.g. reviewing Monday's content on Tuesday), the biggest pay-offs will come from delaying in-lecture review by a week or more, when long-term retention is the instructional goal.
- Instructors can create homework assignments that require review of previously covered content. Ideally, old and new topics should be mixed together.
- When creating a distributed practice schedule, instructors should consider how long they want students to retain the information. As a rough rule, longer spacing intervals increase the duration of retention. However, spacing intervals can be too long. As quick (but somewhat rough) guidelines, here are some suggested spacing intervals for a few retention intervals:
  - One-day spacing is good for 1 week of retention.
  - One-week spacing is good for 2 months of retention.
  - One-month spacing is good for 1 year of retention.
- Many of the methods we suggest for retrieval practice can (and should) be implemented on a distributed schedule. For example, low-stakes quizzes at the start of a lesson can act as a form of spaced retrieval of important topics from previous lessons. Cumulative exams also provide spaced retrieval practice.

References

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

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861–876.

Agarwal, P. K., D'Antonio, L., Roediger, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3, 131–139.

Arnold, K. M. & McDermott, K. B. (2013a). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review*, 20, 507–513.

Arnold, K. M. & McDermott, K. B. (2013b). Test-potentiated learning: distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 39, 940–945.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science,* 4, 316–321.

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.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.

Bird, S. (2011). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, 32, 435–452.

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

Blunt, J. R. & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106, 849–858.

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. & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514–527.

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.

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, 20, 941–956.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13, 273–281.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 918–928.

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.

Callender, A. A. & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, 34, 30–41.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1563–1569.

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

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: review of recent research and implications for instruction. *Educational Psychology Review*, 24, 369–378.

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.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: theoretical analysis and practical implications. *Experimental Psychology*, 56, 236–246.

Crowder, R. G. (1976). *Principles of Learning and Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. In: B. Ross (Ed.), *The Psychology of Learning and Motivation, Volume 53* (pp. 63–148). Burlington, VT: Academic Press.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, 79, 162–170.

Ebbinghaus, H. (1885/1964). *Memory: A Contribution to Experimental Psychology* (translated by H. A. Ruger & G. E. Bussenius, 1913). New York: Dover.

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

Fisher, R. P. & Geiselman, R. E. (1992). *Memory Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL: Charles C. Thomas.

Gerbier, E. & Toppino, T. C. (2015). The effect of distributed practice: neuroscience, cognition, and education. *Trends in Neuroscience and Education*, 4, 49–59.

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.

Hartwig, M. K. & Dunlosky, J. (2012). Study strategies of college students: are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19, 126–134.

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.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In: R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium* (pp. 77–97). Potomac, MD: Lawrence Erlbaum Associates.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, 18, 879–919.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19, 528–558.

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, 21, 1544–1550.

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: long-term benefits for factual and higher-level learning. *Learning and Instruction*, 36, 38–45.

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. & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 704–719.

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

Karpicke, J. D. & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772–775.

Karpicke, J. D. & Grimaldi, P. J. (2012). Retrieval-based learning: a perspective for enhancing meaningful learning. *Educational Psychology Review*, 24, 401–418.

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

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

Kornell, N. (2009). Optimising learning using flashcards: spacing is more effective than cramming. *Applied Cognitive Psychology*, 23, 1297–1317.

Kornell, N. & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, 14, 219–224.

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

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

Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect: strong effects on weak theoretical grounds. *Zeitschrift für Psychologie*, 222, 71–81.

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: enhancing students' memory for vocabulary. *Instructional Science*, 42, 373–388.

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

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological Science*, 25, 639–647.

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, 23, 1337–1344.

Logan, J. M. & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition*, 15, 257–280.

Lyle, K. B. & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, 38, 94–97.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494–513.

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

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, 20, 3–21.

Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, 14, 194–199.

Marsh, E. J., Lozito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors made on a multiple-choice test. *Memory*, 20, 645–653.

Mawhinney, V. T., Bostow, D. E., Laws, D. R., Blumenfeld, G. J., & Hopkins, B. L. (1971). A comparison of students' studying behavior produced by daily, weekly, and three-week testing schedules. *Journal of Applied Behavior Analysis*, 4, 257–264.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, 37, 1077–1087.

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.

Novak, J. D. (1990). Concept mapping: a useful tool for science education. *Journal of Research in Science Teaching*, 27, 937–949.

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.

Putnam, A. L. & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition*, 41, 36–48.

Pyc, M. A. & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60, 437–447.

Pyc, M. A. & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science*, 330, 335.

Rawson, K. A. & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, 97, 70–80.

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

Roediger, H. L. (2013). Applying cognitive psychology to education: translational educational science. *Psychological Science in the Public Interest*, 14, 1–3.

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

Roediger, H. L. & Karpicke, J. D. (2006b). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210.

Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: best practices and boundary conditions. In: G. M. Davies & D. B. Wright (Eds), *Current Issues in Applied Memory Research* (pp. 13–49). Brighton, UK: Psychology Press.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011a). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17, 382–395.

Roediger, H. L., Putnam, A. L., & Smith, M. (2011b). 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. Volume 55* (pp. 1–36). Oxford: Elsevier.

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, 40, 4–17.

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*, 27, 635–643.

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

Schacter, D. L. & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology*, 1, 60–71.

Smith, M. A. & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, 22, 784–802.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1712–1725.

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. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, 73, 99–115.

Son, L. K. & Kornell, N. (2008). Research on the allocation of study time: key studies from 1890 to the present (and beyond). In: J. Dunlosky & R. A. Bjork (Eds), *A Handbook of Memory and Metamemory* (pp. 333–351). Hillsdale, NJ: Psychology Press.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 6313–6317.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3, 161–164.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Özsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology*, 22, 685–695.

Willingham, D. B. (2009). *Why Don't Students Like School?* San Francisco, CA: Jossey-Bass.

Zaromb, F. M. & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38, 995–1008.