Recent Research on Human Learning Challenges Conventional Instructional Strategies

Doug Rohrer and Harold Pashler

EDUCATIONAL RESEARCHER 2010 39: 406
DOI: 10.3102/0013189X10374770

Recent Research on Human Learning Challenges Conventional Instructional Strategies

Doug Rohrer and Harold Pashler

There has been a recent upsurge of interest in exploring how choices of methods and timing of instruction affect the rate and persistence of learning. The authors review three lines of experimentation—all conducted using educationally relevant materials and time intervals—that call into question important aspects of common instructional practices. First, research reveals that testing, although typically used merely as an assessment device, directly potentiates learning and does so more effectively than other modes of study. Second, recent analysis of the temporal dynamics of learning show that learning is most durable when study time is distributed over much greater periods of time than is customary in educational settings. Third, the interleaving of different types of practice problems (which is quite rare in math and science texts) markedly improves learning. The authors conclude by discussing the frequently observed dissociation between people's perceptions of which learning procedures are most effective and which procedures actually promote durable learning.

Keywords: cognition; instructional design/development; memory

The experimental study of human learning and memory began more than 100 years ago and has developed into a major enterprise in behavioral science. Although this work has revealed some striking laboratory phenomena and elegant quantitative principles, it is disappointing that it has not thus far given teachers, learners, and curriculum designers much in the way of concrete and nonobvious advice that they can use to make learning more efficient and durable. In the past several years, however, there has been a new burst of effort by researchers to identify and test concrete principles that have this potential, yielding a slew of recommended strategies that have been listed in recent reports (e.g., Halpern, 2008; Mayer, in press; Pashler et al., 2007). Some of the most promising results involve the effects of testing on learning and different ways of scheduling study events. Those skeptical of behavioral research might assume that principles of learning would already be fairly obvious to anyone who has been a student, yet the results of recent experimentation challenge some of the most widely used study practices. We discuss three topics, focusing on the effects of testing, the role of temporal spacing, and the effects of interleaving different types of materials.

Learning Through Testing

Tests of student mastery of content material are customarily viewed as assessment devices, used to provide incentives for students (and in some cases teachers and school systems as well). However, memory research going back some years has revealed that a test that requires a learner to retrieve some piece of information can directly strengthen the memory representation of this information (e.g., Spitzer, 1939). More recent studies, however, have shown that a combination of study and tests is more effective than spending the same amount of time reviewing the material in some other way, such as rereading it (e.g., Carrier & Pashler, 1992; Cull, 2000; for reviews, see McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006b). Interestingly, however, surveys of college students show that most of them study almost entirely by rereading, with self-testing relatively rarely employed (Carrier, 2003; Karpicke, Butler, & Roediger, 2009).

Recent research shows that testing not only enhances learning but also slows the rate of forgetting. Roediger and Karpicke (2006a) found that a period of time devoted to a combination of study and tests rather than study alone impaired performance on a test given 5 minutes later yet improved performance on a test given 1 week later (Figure 1). Testing also retarded the rate of forgetting in two studies with test delays as long as 42 days (Carpenter, Pashler, Wixted, & Vul, 2008).

While one might attribute the benefit of the initial test to heightened attention, it seems more likely that it arises from the retrieval itself, as evidenced by a study by Kang, McDermott, and Roediger (2007) showing that an initial test requiring respondents to choose the correct answer from a list of alternatives (i.e., a multiple-choice question) did not produce as much benefit as a test requiring recall (i.e., a short-answer question). Moreover, these authors found that explicit retrieval, as required by a recall task rather than a recognition task, strengthened knowledge better than a multiple-choice test even when the final test itself involves multiple choice—and thus the effect is not attributable to a simple principle that practicing a given type of test best enhances performance on the same type of test.

A number of studies have shown that sizable benefits of testing generalize to classroom-based studies. In a study reported by

Educational Researcher, Vol. 39, No. 5, pp. 406–412
DOI: 10.3102/0013189X10374770

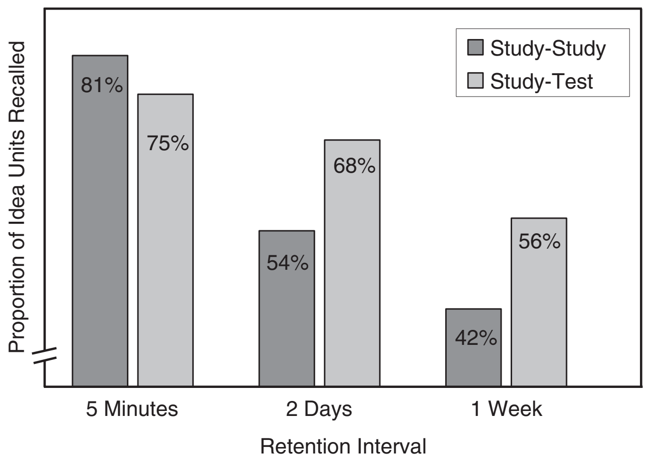

FIGURE 1. *Testing retards the rate of forgetting. In this study by Roediger and Karpicke (2006a), college students read a passage and then, after a 2-minute delay, either reread the passage (the study-study condition) or wrote as much of the information as they could recall (study-test). Respondents were given a final test after a test delay (or "retention interval") of 5 minutes, 2 days, or 1 week. The initial test depressed final test scores after a 5-minute delay (*d *= 0.52) yet improved final test scores after a delay of 2 days (*d *= 0.95) or 1 week (*d *= 0.83). Error bars represent standard errors of the mean. Adapted from Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention.* Psychological Science, 17, *249–255.*

Butler and Roediger (2007), students viewed three video lectures on the topic of art history. Each one was followed by a short-answer test, a multiple-choice test, or a summary review of the lecture. One month later, a final test was given. Retention was substantially better for the items included in the short-answer test; the other two conditions were not reliably different. Similarly, McDaniel, Anderson, Derbish, and Morrisette (2007) conducted an experiment in conjunction with a college course and found that students' final test performances were improved if their chapter readings were followed by review questions (e.g., "All preganglionic axons, whether sympathetic or parasympathetic, release ______ as a neurotransmitter"; McDaniel et al., 2007, p. 499) rather than statements ("All preganglionic axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter"; p. 499). Finally, in a study reported by Carpenter, Pashler, and Cepeda (2009), as well as a series of ongoing experiments by researchers at Washington University (e.g., Agarwal, Roediger, McDaniel, & McDermott, 2008), middle school students better recalled the material in their science, social science, or history class (which was taught by their regular teachers) if the classroom presentations were followed by review questions (with answer feedback). Test delays in both studies were as long as 9 months.

The testing effect also has been demonstrated with a variety of tasks and materials. For instance, testing has been shown to improve foreign vocabulary learning (Carrier & Pashler, 1992; Karpicke & Roediger, 2008), retention of content from passages and scientific articles (e.g., Kang et al., 2007; Roediger & Karpicke, 2006a), and map learning (Carpenter & Pashler, 2007; Rohrer, Taylor, & Sholar, 2010). Likewise, testing effects have been observed in a variety of learning environments, including self-paced study outside the classroom (McDaniel et al., 2007), instructor-paced instruction (e.g., Carpenter et al., 2009), and multimedia learning (Johnson & Mayer, 2009). In brief, scaled-up efficacy studies have not yet been done; researchers have yet to identify a boundary condition of the testing effect.

The results described above suggest that instructional practices would be more effective if the proportion of learning time that learners spend retrieving information were dramatically increased. Some of the studies discussed here have shown that even simple self-testing methods, such as having students retrieve everything they can recall from a text or lecture, can be more effective than the most commonly used study strategies. However, it seems plausible that technological refinements of these strategies could further enhance the benefits of testing and retrieval.

Spacing of Practice: Temporal Variables

In conventional classroom instruction, it is common for students to encounter materials only over a fairly short time period (e.g., weekly vocabulary lists), and at all levels of education, programs that try to compress learning into very short time spans seem to be ever popular (e.g., summer boot camps in biomedical techniques; immersion programs in foreign languages). Is it sensible to compress instruction in this manner?

Whether a particular set of study materials should be *massed* into a single presentation or *distributed* across multiple sessions has been a key question in learning research for more than a century. Almost invariably, the data show that, if a given amount of study time is distributed or *spaced* across multiple sessions rather than massed into a single session, performance on a delayed final test is improved—a finding known as the *spacing effect.* These results have led many authors to conclude that teachers and curriculum designers should rely more heavily on spacing (e.g., Bahrick, 1979; Bjork, 1979; Dempster, 1987, 1988, 1989). Yet few educators have heeded this advice, as evidenced, for instance, by a glance at students' textbooks.

One possible reason for this continued neglect is the generally poor ecological validity of most spacing experiments. There are notable exceptions, though. For instance, whereas the vast majority of spacing studies allow participants to finish an experiment in less than an hour, Harry Bahrick and his colleagues employ test delays lasting several years. In a study by Bahrick and Phelps (1987), for instance, spacing boosted participants' recall of Spanish vocabulary learned in a college course taken 8 years earlier, and Bahrick, Bahrick, Bahrick, and Bahrick (1993) found that spacing boosted the *authors'* retention of foreign language vocabulary after a test delay of 5 years. Thus the spacing effect appears to hold over educationally relevant time periods.

But other doubts about the applicability of the spacing effect have received little attention until recently. For example, spacing studies with cognitive tasks traditionally have relied on very simple tasks (e.g., Spanish–English pairs), but recent studies have demonstrated sizable spacing effects with more complex forms of learning. For instance, Moulton et al. (2006) found that 1-week spacing enhanced surgical performance assessed 1 month later. Similarly, two studies of mathematics learning (Rohrer & Taylor, 2006, 2007) found that college students who spaced rather than massed their

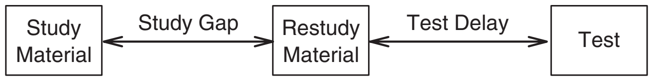

FIGURE 2. *Procedure of typical spacing experiment. Study events are separated by a varying study gap, and the test delay between the final study event and the test is either fixed or manipulated.*

practice of a combinatoric procedure performed dramatically better on a subsequent test consisting of novel problems of the same kind. Finally, in a study by Bird (in press), adults who were learning English were taught to identify and correct subtle grammatical errors (e.g., "When have you arrived?" should be "When did you arrive?"), and a greater degree of temporal distribution improved performance on a subsequent test given 60 days after the final learning session. Thus although there may exist tasks for which spacing is not helpful, extant data suggest that spacing improves long-term learning of at least some kinds of abstract learning.

A final doubt about the ecological validity of the spacing effect stems from the paucity of experiments conducted in classroom settings, but this shortcoming also has been addressed recently. Seabrook, Brown, and Solity (2005) found that 5-year-old children benefited from a greater temporal distribution of phonics lessons. In a study by Metcalfe, Kornell, and Son (2007), a 6-week program that utilized spacing and testing (vs. the appropriate control) boosted the performance of middle-school students. Finally, Carpenter et al. (2009) found that a greater degree of spacing boosted eighth-grade students' recall of the material in their U.S. history course when tested 9 months after the final exposure.

Interestingly, there are several older classroom-based studies often cited as demonstrations of the spacing effect that are, upon closer inspection, not studies of spacing at all—a point their authors seem to have been aware of. For instance, in a study by Smith and Rothkopf (1984), four tutorials were massed or distributed across 4 days, but each tutorial concerned a different topic, which means that the design assessed the effect of rest breaks (e.g., one long lecture each week vs. three short lectures) rather than the effect of spacing. Likewise, in an experiment reported by Gay (1973), two study sessions were spaced either 1 or 14 days apart, but both groups were tested 21 days after the first session, which means that students with the longer spacing interval benefited from a much shorter delay between their final learning session and the test. Thus the more recent studies cited earlier in this article addressed a gap in the literature.

In addition to issues of ecological validity, recent spacing studies have sought to answer critical questions about how the spacing effect can best be exploited. For example, does the spacing effect depend on the time interval over which the material is distributed, and if so, how? This question has been addressed by examining a rather simple case: the effect of studying the same piece of information on two separate occasions, separated by a specified study gap (Figure 2). Memory is assessed after a final *test delay* (measured from the second study event). One might suppose that, when the test delay is fixed, an increase in the study gap could only impair memory because it increases the time over which forgetting of the first study event could occur. However, increases in gap often lead to better test performance, with some fall-off occurring as the gap is increased beyond a certain point (Crowder, 1976; Glenberg & Lehmann, 1980).

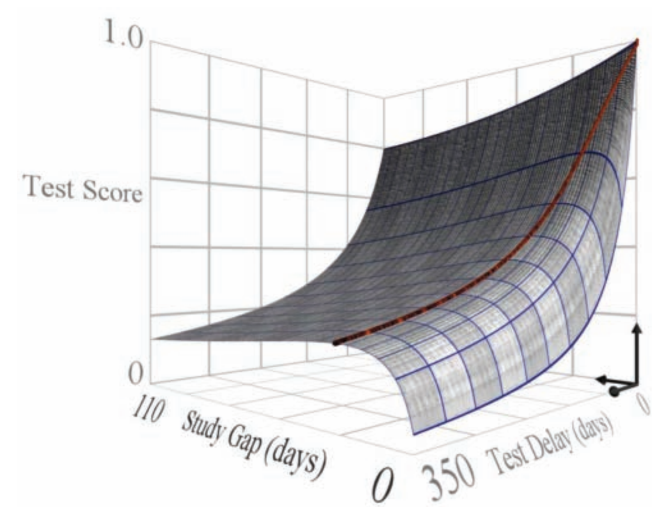

FIGURE 3. *Optimal spacing gap as a function of test delay. In this study by Cepeda, Vul, Rohrer, Wixted, and Pashler (2008), two study events were separated by a varying study gap (0–105 days) and followed by a varying test delay (7–350 days). The data were well fit (*R^2 *= .98) by the illustrated surface, which captures four notable features of the data: (a) For each study gap, test score decreases with negative acceleration as test delay increases; (b) for any nonzero test delay, test score is a nonmonotonic function of study gap; (c) as test delay increases, the optimal study gap increases, as indicated by the direction of the heavy line (red in the online journal) along the ridge of the surface; and (d) as test delay increases, the optimal study gap represents a declining proportion of test delay.*

Until recently, most research on this issue looked at short time delays (with some notable exceptions by Harry Bahrick and his colleagues, as mentioned above), making advice about how to exploit the effect rather speculative and inconsistent. Recent studies have begun exploring the interaction of the study gap and test delay systematically, using realistic (if simple) educational materials and time delays long enough to be relevant to practical situations. For example, in studies reported by Glenberg and Lehmann (1980) and Cepeda et al. (2009), a final test was given 7 to 10 days after the second study event, and a 1-day gap produced better performance than a shorter or longer gap. However, Cepeda et al. also found that, when test delay was lengthened to 6 months, test performance improved dramatically as the gap was increased from a few minutes to about 1 month, with a shallow drop in test scores occurring beyond for longer study gaps. Evidently, retention is maximized when the gap is some small fixed ratio of the final test delay.

This point has been explored further in a larger and more systematic study using a wide range of gaps and test delays. In a study reported by Cepeda, Vul, Rohrer, Wixted, and Pashler (2008), 1,354 participants studied unfamiliar facts on two separate occasions, separated by gaps ranging from 20 minutes to 105 days, followed by a test delay of between 7 and 350 days. The results revealed a surface with a saddle, showing that the optimal study gap tracked the test delay (Figure 3). The optimal gap seems to be a slowly declining proportion of the test delay, but for practical purposes, a gap of approximately 5% to 10% of the test delay is optimal.

This finding—optimal study gap increases with the duration of the test delay—yields rather concrete advice: If one wishes to retain information for a long period of time, the interval of time over which one studies or practices should be moderately long as well. For instance, if the goal is very long-term retention, or even lifelong retention, which is presumably the aim in most educational contexts, then previously studied material should be revisited at least a year after the first exposure—something that happens rather rarely in most educational systems. In brief, sufficiently long spacing gaps have the potential to improve long-term retention.

Interleaving

If multiple kinds of skills must be learned, the opportunities to practice each skill may be ordered in two very different ways: *blocked* by type (e.g., aaabbbccc) or *interleaved* (e.g., abcbcacab). Until recently, experimental comparisons of blocked and interleaved practice had been limited to studies of motor skill learning, where it has been found that interleaving increases learning (Carson & Wiegand, 1979; Hall, Domingues, & Cavazos, 1994; Landin, Hebert, & Fairweather, 1993; Shea & Morgan, 1979). For example, when baseball players practiced hitting three types of pitches (e.g., curve ball) that were either blocked or interleaved, interleaving improved performance on a subsequent test in which the batters did not know the type of pitch in advance—as would be the case in a real game, of course (Hall et al., 1994).

Recent experiments have shown that interleaving can enhance the learning of cognitive skills as well. For example, in one recent study by Kornell and Bjork (2008), adults viewed numerous paintings by each of 12 artists with similar styles, with the paintings either blocked by artist or interleaved. The interleaving strategy improved performance on a subsequent test in which the respondents were shown previously unseen paintings by the same artists and asked to identify the artist. In essence, interleaving improved their ability to discriminate between the different kinds of styles.

If interleaving improves discriminability, it would seem that interleaving might have particularly large effects on mathematics learning. This is because mathematics proficiency requires the ability to choose the appropriate solution or method for a given kind of problem, and superficially similar kinds of problems often demand different kinds of solutions. For instance, the integration problems $\int ex^e dx$ and $\int xe^x dx$ resemble each other yet require different techniques (the latter requires integration by parts). Likewise, if a statistics course is to prove useful, students must be able to choose the appropriate statistical test for a given kind of research design. Moreover, being able to identify the appropriate solution is arguably more important than knowing how to execute the solution, because once the appropriate method is identified (e.g., Mann-Whitney *U* statistical test), it can often be executed by a computer program.

Recent studies confirm that interleaving can have powerful benefits in math learning. In one study reported by Rohrer and Taylor (2007), college students learned to find the volume of four obscure solids by solving practice problems that were interleaved or blocked, and interleaving boosted subsequent test scores by a factor of three ($d = 1.34$). Furthermore, although one might reasonably suspect that the benefit of interleaving is merely another instance of the spacing effect discussed above, because interleaving inherently ensures a greater degree of spacing than does blocking (e.g., abcbcacab vs. aaabbbccc), a recent study with young children found a large benefit of interleaving ($d = 1.21$) even though the degree of spacing was held fixed (Taylor & Rohrer, in press).

Despite the empirical support for interleaving, virtually all mathematics textbooks rely primarily on blocked practice, as each section is followed by a set of practice problems devoted to the material in that same section. Consequently, students must solve several problems of the same kind in immediate succession—a degree of repetition that has been shown to produce dramatically diminishing returns (Rohrer & Taylor, 2006). More important, though, blocked practice ensures that students know the appropriate technique or relevant concept *before* they read the problem. In some cases, in fact, students can solve word problems without reading the words; they simply pick out the numerical data and repeat the procedure used in the previous problems. Thus blocked practice provides students with a crutch that is unavailable during cumulative final exams and standardized tests and in the real-world situations for which they are presumably being trained. It is not surprising if they often struggle when asked to demonstrate a skill they have not previously practiced.

Fortunately, interleaved practice is fairly easily incorporated in textbooks in mathematics (or physics or chemistry). The practice problems in the text need merely to be rearranged, and the lessons remain unchanged. Thus each section would be followed by the usual number of practice problems, but most of these problems would be based on material from previous sections. Besides allowing problems of different kinds to be interleaved, this format provides a spacing effect because the problems on any given topic are distributed throughout the textbook. Sometimes known as mixed review or cumulative review, it proved superior to blocked practice in a recent classroom-based study (Mayfield & Chase, 2002).

Why Are Inferior Strategies So Popular?

The experimental findings in favor of testing, spacing, and interleaving raise an obvious question: Why don't people notice the benefits of these methods and spontaneously adopt them? The blame cannot be placed on the logistical costs of these learning strategies, which generally require no more time and resources than do the alternative strategies—indeed, spacing and interleaving require only a change in the scheduling of study events or practice problems. Rather, the underutilization of these learning strategies appears to reflect the widespread (but erroneous) feeling that these strategies are less effective than their alternatives.

Interestingly, learners often fail to recognize the efficacy of testing, spacing, and interleaving even when they have just used these strategies along with others. For instance, in a study by Karpicke and Roediger (2008) showing that the recall of Swahili vocabulary words was nearly tripled if respondents relied on a learning strategy that emphasized testing (e.g., *mashua*–?) rather than study (*mashua–boat*), respondents' predictions of their final recall performances were about equal for the two strategies, even though respondents gave these predictions immediately *after* they had used both learning strategies. Likewise, in the aforementioned study by Kornell and Bjork (2008) showing that respondents' ability to identify the artist who created a painting they had not seen during training was greater if the study of the artist's other paintings had been interleaved with other artists' paintings

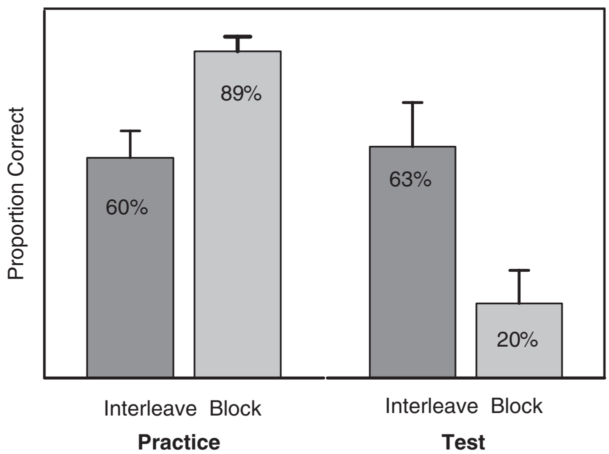

FIGURE 4. *Reversal of efficacy. In a study by Rohrer and Taylor (2007), college students learned to solve four kinds of mathematics problems. All respondents received the same practice problems and saw the complete solution immediately after each practice problem. The four kinds of practice problems were either interleaved or blocked by kind. Interleaving impaired practice session performance (*d *= 1.06) yet improved scores on a final test (*d *= 1.34) given 1 week after the final practice problem. Error bars represent standard errors of the mean.*

rather than blocked by artist, 78% of the respondents predicted that the blocked order would be superior even though the query was given after the experiment was complete.

This failure to recognize the superiority of testing, spacing, or interleaving over alternatives may reflect the fact that these strategies, although beneficial to subsequent test performance, tend to produce more errors during the learning session (Schmidt & Bjork, 1992). For instance, in the interleaving study by Rohrer and Taylor (2007), interleaving improved final test performance yet *reduced* accuracy during the practice sessions (Figure 4). A similar reversal is sometimes caused by spacing (e.g., Bregman, 1967; Cepeda et al., 2009; Landauer & Bjork, 1978). In brief, if people tend to judge the efficacy of a learning strategy on the basis of their performance during training, they will choose strategies that sometimes yield suboptimal long-term learning. For this reason, strategies such as interleaving and spacing, which impair training performance yet improve performance on a delayed test, have been dubbed "desirable difficulties" by Robert Bjork and his colleagues (e.g., Schmidt & Bjork, 1992).

Research on people's ability to assess their own learning—a skill of *metacognition*—also reveals some additional reasons (beyond those mentioned above) why frequent testing is likely to be essential in effective learning procedures. When people reread material that they have studied, and attempt to estimate how well they know individual facts, their estimates show little correlation with subsequent performance on a test (e.g., Dunlosky & Lipko, 2007; Koriat & Bjork, 2006). By contrast, when they are tested, their estimates become much more accurate (Jang & Nelson, 2005; Rawson & Dunlosky, 2007). Other manipulations can also enhance estimation accuracy (Koriat & Bjork, 2006; Thiede, Dunlosky, Griffin, & Wiley, 2005). Thus a student who uses the most common strategy of rereading a textbook will have little basis for determining which material needs more study and which does not (Metcalfe & Finn, 2008; Son & Metcalfe, 2000). By contrast, when self-testing is used extensively, learners become well informed as to which content needs additional study and which does not.

Conclusion

There is much interest at present in the overall goal of turning education into a more evidence-based enterprise. Naturally, this will require contributions from many different fields. As the present review discloses, experimental research on human learning and memory is beginning to make a distinctive contribution to this enterprise, offering concrete advice about how the mechanics of instruction can be optimized to enhance the rate of learning and maximize retention. In some cases, the results reveal that study and teaching strategies that do not require any extra time can produce two- or threefold increases in delayed measures of learning. It is striking how often these strategies differ from conventional instructional and study methods. If educational practices (ranging from textbook layout and educational software design to the study and teaching strategies used by students and teachers) are adjusted to exploit the kinds of findings discussed here, it ought to be possible to significantly enhance educational and training outcomes.

NOTE

This work was supported by a collaborative activity grant from the James S. McDonnell Foundation; by the Institute of Education Sciences, U.S. Department of Education (Grant No. R305B070537); and by the National Science Foundation (Grant No. BCS-0720375 to H. Pashler and Grant No. SBE-582 0542013 to the University of California, San Diego, Temporal Dynamics of Learning Center). The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, the U.S. Department of Education, or other granting agencies that have supported this work.

REFERENCES

Agarwal, P. K., Roediger, H. L., McDaniel, M. A., & McDermott, K. B. (2008, May). *Improving student learning through the use of classroom quizzes*. Poster presented at the 20th Annual Meeting of the Association for Psychological Science, Chicago, IL.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*, 296–308.

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.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over eight years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344–349.

Bird, S. (in press). Effects of distributed practice on the acquisition of L2 English syntax. *Applied Psycholinguistics*.

Bjork, R. A. (1979). Information-processing analysis of college teaching. *Educational Psychologist*, *14*, 15–23.

Bregman, A. S. (1967). Distribution of practice and between-trials interference. *Canadian Journal of Psychology*, *21*, 1–14.

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.

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

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., Pashler, H., Wixted, J., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory and Cognition*, *36*, 438–448.

Carrier, L. M. (2003). College students' choices of study strategies. *Perceptual and Motor Skills*, *96*, 54–56.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, *20*, 632–642.

Carson, L. M., & Wiegand, R. L. (1979). Motor schema formation and retention in young children: A test of Schmidt's schema theory. *Journal of Motor Behavior*, *11*, 247–251.

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

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*, *11*, 1095–1102.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235.

Dempster, F. N. (1987). Time and the production of classroom learning: Discerning implications from basic research. *Educational Psychologist*, *22*, 1–21.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*, 627–634.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*, 309–330.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16,* 228–232.

Gay, L. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology*, *64*, 171–182.

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory and Cognition*, *8*, 528–538.

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

Halpern, D. F. (2008, March). *25 learning principles to guide pedagogy and the design of learning environments*. Handout distributed at the keynote address at the Bowling Green State University Teaching and Learning Fair, Bowling Green, OH.

Jang, Y., & Nelson, T. O. (2005). How many dimensions underlie judgments of learning and recall? Evidence from state-trace methodology. *Journal of Experimental Psychology: General*, *134,* 308–326.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*, 621–629.

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

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

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

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory and Cognition*, *34,* 959–972.

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

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Skykes (Eds.), *Practical aspects of memory* (pp. 625–632). London, UK: Academic Press.

Landin, D. K., Hebert, E. P., & Fairweather, M. (1993). The effects of variable practice on the performance of a basketball skill. *Research Quarterly for Exercise and Sport*, *64*, 232–236.

Mayer, R. E. (in press). *Applying the science of learning*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, *35*, 105–123.

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., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review*, *14*, 200–206.

Metcalfe, J. & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review*, *15,* 174–179.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology*, *19*, 743–768.

Moulton, J., Mundy, K., Welmond, M., & Williams, J. (2006). Educational reforms in sub-Saharan Africa: Paradigm lost. *Africa Today*, *52,* 141–144.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning. IES practice guide* (NCER 2007–2004). Washington, DC: National Center for Education Research. Retrieved February 28, 2008, from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072004.pdf

Rawson, K., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*, 559–579.

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *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.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209–1224.

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

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

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, *19*, 107–122.

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

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*, 341–358.

Son, L. K. & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory,* & *Cognition*, *26,* 204–221.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30,* 641–656.

Thiede, K. W., Dunlosky, J., Griffin, T. & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31,* 1267–1280.

Taylor, K., & Rohrer, D. (in press). The effect of interleaving practice. *Applied Cognitive Psychology*.

AUTHORS

DOUG ROHRER is a professor of psychology at the University of South Florida, Psychology, PCD4118G, 4202 East Fowler Avenue, Tampa, FL 33620; *drohrer@usf.edu*. His research focuses on learning, with a special emphasis on mathematics learning.

HAROLD PASHLER is a professor of psychology at the University of California, San Diego, Department of Psychology 0109, 9500 Gilman Drive, La Jolla, CA 92093; *hpashler@ucsd.edu*. His research focuses on memory, human learning, and attention.

Manuscript received February 8, 2010
Revisions received April 1, 2010
Accepted April 14, 2010