The Effects of Spacing and Mixing Practice Problems

Doug Rohrer
University of South Florida

Sets of mathematics problems are generally arranged in 1 of 2 ways. With *blocked practice*, all problems are drawn from the preceding lesson. With *mixed review*, students encounter a mixture of problems drawn from different lessons. Mixed review has 2 features that distinguish it from blocked practice: Practice problems on the same topic are distributed, or spaced, across many practice sets; and problems on different topics are intermixed within each practice set. A review of the relevant experimental data finds that each feature typically boosts subsequent performance, often by large amounts, although for different reasons. Spacing provides review that improves long-term retention, and mixing improves students' ability to pair a problem with the appropriate concept or procedure. Hence, although mixed review is more demanding than blocked practice, because students cannot assume that every problem is based on the immediately preceding lesson, the apparent benefits of mixed review suggest that this easily adopted strategy is underused.

Key words: Distributed; Mathematics; Mixed; Practice; Spacing; Review

Most mathematics students devote a large fraction of their study time to practice problems, and yet the effect of this practice on proficiency receives comparatively little attention from researchers. Although practice problems are characterized by many features, this commentary focuses on just one easily manipulated feature: the order in which practice problems are arranged. Although such a manipulation might seem trivial, the data show that merely reordering practice problems can dramatically affect subsequent performance.

TWO WAYS OF ARRANGING PRACTICE PROBLEMS

In virtually all mathematics textbooks, each lesson is followed by a set of practice problems, a *practice set*, typically arranged in one of two fundamentally different ways. In *blocked practice*, a group of consecutive problems is devoted to the immediately preceding lesson. A lesson on ratios, for example, might be followed by one or two dozen ratio problems. Hence, although the problems might include a combination of procedural problems, word problems, and so forth, the

This work was supported by the Institute of Education Sciences in the U.S. Department of Education (Grants R305H020061 and R305H040108). The opinions expressed are those of the author and do not necessarily represent the views of the Institute of Education Sciences.

underlying topic would be the same. In an alternative arrangement known as *mixed review*, a group of problems is drawn from many different lessons and ordered so that problems requiring the same skill or concept do not appear consecutively. Thus, mixed review has two defining components: (a) problems on a particular topic are distributed across many practice sets (which yields review), and (b) problems relating to different topics are mixed within each practice set. Thus, spaced practice and mixed practice are necessary components of mixed review, and it follows that neither term is a synonym for mixed review. Incidentally, mixed review is sometimes described as "cumulative review," but I avoid that term because it is often used to describe *unmixed* review (e.g., a few problems regarding Lesson 6–1, followed by problems regarding Lesson 6–2, etc.).

Mathematics textbooks typically include both blocked practice and mixed review, but most rely far more heavily on blocked practice. For example, in a middle school textbook by Glencoe (2001), each lesson is followed by two or three dozen problems of blocked practice, followed by a separate set of three to five mixed review problems. Similarly, in the elementary school Everyday Math series (University of Chicago School Mathematics Project, 2007), each set of blocked practice problems is followed by a smaller set of mixed review problems. In contrast, the Saxon textbook series for Grades K to 12 (e.g., Saxon, 1997) appears to be unique in relying more heavily on mixed review than on blocked practice.

Mixed review can be accomplished in a number of ways. For example, the nth lesson could be followed by set of problems drawn more heavily from recent lessons, as would occur if a practice set included the following problems (but not in the order shown):

- Six problems on Lesson n (including procedural problems, word problems, etc.)
- Three problems on Lesson $(n - 1)$
- Two problems on Lesson $(n - 2)$
- One problem each on Lessons $(n - 3)$, $(n - 4)$, $(n - 5)$, $(n - 10)$, $(n - 20)$, $(n - 40)$, $(n - 60)$, and $(n - 90)$.

To ensure a mixture, these problems would then be ordered so that no two problems from the same lesson appeared consecutively. In cases in which $n - k < 1$, where $k \in \{1, 2, 3, 4, 5, 10, 20, 40, 60, 90\}$, as occurs, for example, when the algorithm calls for a problem on Lesson $(n - 10)$ to be included in the fifth practice set, the needed problem can be replaced by one relating to either a more recent lesson or a lesson from a prior year. In brief, mixed review differs from blocked practice in salient ways, and the relative merits of these alternative formats must be assessed if students are to maximize the effectiveness and efficiency of their efforts.

EXPERIMENTAL DATA

Although a number of studies have assessed various features of mixed review, Mayfield and Chase (2002) conducted what is apparently the only explicit comparison of mixed review and blocked practice. In their experiment, college students in a remedial program attended numerous sessions over several months in which they learned several rules for operations with exponents (e.g., $ax^m \cdot bx^n = abx^{m+n}$). Although every student was taught the rules in the same manner, each student was randomly assigned to one of three practice schedules, two of which are relevant. For the blocked practice group, each practice session was devoted to a single rule. For the mixed review group, each practice session included an interleaving of problems for each of the previously learned rules, which provided mixing and a greater degree of spacing than the blocked practice group saw. The last practice session was followed by two tests: the first after a delay of 1 or 2 days, and a second one 4 to 12 weeks later. The mixed-review group outscored the blocked practice group on both delayed tests, although the difference after the longer delay was not statistically significant. The interpretation of this result is slightly complicated, though, because the mixed review and blocked practice conditions differed in several ways other than the degree to which practice problems were spaced and mixed. For instance, at least a third of the practice problems were not common to both groups, and the proportion of problems devoted to each algebraic rule differed across groups. Still, to the extent that those confounding variables did not favor mixed review, and there is no apparent reason to believe either did so, this finding provides support for mixed review.

Whereas Mayfield and Chase (2002) informatively assessed the *conjoint* effects of spacing and mixing practice, an independent assessment of each feature can provide a better understanding of how each exerts its effects. Toward this aim, I review the literature on the independent effects of three variables: heavy repetition, which is a common (but not defining) feature of blocked practice, and the two defining features of mixed review (spacing and mixing). I address three questions:

1. With regard to heavy repetition, once students have solved one kind of practice problem on a topic, is there a benefit in doing several more of those problems immediately afterward?
2. What is the effect of distributing problems on one topic across multiple practice sets?
3. When a practice set includes problems on several topics, what is the effect of mixing problems on different topics?

Effect of Heavy Repetition

Although a blocked practice set might include a variety of procedural problems, word problems, and so forth, the sheer number of problems devoted to the same topic means that blocked practice sets often include multiple problems of the same kind (e.g., a half dozen problems requiring students to find the least common multiple of two positive integers). Such repetition is not a necessary characteristic of blocked practice—the practice set might consist of several dozen problems of different kinds on the same topic—but heavy repetition is nevertheless a common characteristic of blocked practice that distinguishes it from mixed review. Thus, to assess the utility of mixed review, which lacks heavy repetition, it is worth knowing whether heavy repetition pays dividends.

Once a student has correctly solved a particular kind of problem (e.g., finding the least common multiple), immediately working additional problems of the same kind on the same topic constitutes what is known as an *overlearning* strategy. Overlearning is almost uniformly endorsed by those who have written about it. For instance, Hall (1989) wrote that overlearning "will prevent significant losses in retention" (p. 328), and Fitts (1965) concluded that "the importance of [immediately] continuing practice beyond the point in time where some (often arbitrary) criterion is reached cannot be overemphasized" (p. 195). To what extent, though, are these claims supported by data?

Over the last 80 years, studies of overlearning with nonmathematical tasks have shown that test performance is, in fact, improved if learners immediately continue to practice the same task after achieving one success instead of quitting after that first success (for a meta-analysis, see Driskell, Willis, & Copper, 1992), but that endorsement is qualified by two critical caveats. First, the benefit of overlearning dissipates with time. For instance, the largest overlearning effects observed in the Driskell et al. meta-analysis occurred when the time between practice and test was less than a week. Second, each additional unit of effort devoted to overlearning increases test scores by an increasingly smaller amount, and ultimately any additional overlearning has no impact (e.g., Krueger, 1929).

Thus, with these caveats, the results of overlearning studies of nonmathematical tasks suggest that requiring students to work more than a few consecutive mathematics problems of the same type would boost subsequent test scores by only a negligible amount when test delays were nontrivial. This possibility was tested in a recent experiment in which students worked either 3 or 9 practice problems of the same kind in immediate succession (Rohrer & Taylor, 2006). In this study, college students first observed a tutorial describing how to find the number of permutations for a sequence of items with at least one repeated item. For instance, the sequence *abbccc* has 60 permutations. Immediately after the tutorial, the students completed either 3 or 9 practice problems in immediate succession and at a prescribed pace, with each attempt followed immediately by a visually presented solution. The students returned either 1 or 4 weeks later (as determined by random assignment) for a test that included problems of the same kind (without feedback). The additional 6 practice problems provided a negligible benefit after both 1 week (69% vs. 67%) and 4 weeks (28% vs. 27%).

Even if heavy repetition did boost test scores, it would not be advisable unless the benefit was greater than that derived from a different use of the time devoted to the additional repetitions. That is, for practical purposes, the question is not whether 6 additional problems of the same kind provide a benefit but whether the benefit exceeds that of working 6 problems on other topics. In brief, the utility of a learning strategy should be judged not by its effectiveness per se but by its efficiency (or relative effectiveness). The data from this study provide no support for the heavy repetition that is common in blocked practice.

Effect of Spacing

A second salient feature of mixed review is that the practice problems on a topic are distributed across many practice sets, whereas blocked practice entails massing those problems in a single practice set. Note that spacing increases only the temporal distribution of practice problems and not their total number. For example, rather than work 10 problems on the same topic in one session, a student might divide the same 10 problems across two sessions separated by a week. Furthermore, spacing does not entail more recent practice, because in a well-designed spacing experiment, the interval between study and test, the *test delay* (or retention interval), is measured from the last practice problem. With these constraints, numerous studies have found that the spacing of practice across multiple sessions improves subsequent test performance (for a review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

Indeed, the *spacing effect* is arguably one of the largest and most robust findings in learning research, and it appears to have few constraints. It has been demonstrated with a wide range of tasks, including tasks in learning a foreign language (Bloom & Shuell, 1981; Bahrick, Bahrick, Bahrick, & Bahrick, 1993), spelling (Fishman, Keller, & Atkinson, 1968), biology (Reynolds & Glaser, 1964), and mathematics, 1968; (detailed below). Likewise, although most spacing studies involve college students in laboratory settings, spacing has been shown to benefit elementary and middle school students as well (e.g., Fishman et al., 1968; Metcalfe, Kornell, & Son, 2007; Rea & Modigliani, 1985; Seabrook, Brown, & Solity, 2005; Toppino, Kasserman, & Mracek, 1991). Finally, the spacing effect has been shown to hold after test delays of 1 year or more (Bahrick et al., 1993; Cepeda, Vul, Rohrer, Wixted, & Pashler, in press).

Relatively few experiments, however, have assessed the effects of spacing mathematics practice. Moreover, at least two studies that have been cited as instances of a spacing effect in learning mathematics are, in fact, not spacing studies. Smith and Rothkopf (1984) presented four tutorials either in immediate succession or across 4 successive days, but each tutorial concerned a different topic, which means that neither group received spaced practice on any subset of the material. (Still, the results are informative, as the finding suggests, e.g., that a weekly 3-hour college lecture is less effective than three 1-hour lectures.) In Gay (1973), spacing and test delay were conflated. In Experiment 1, for example, learners had either massed practice on Days 1 and 2 (1-day gap) or spaced their practice across Days 1 and 15 (14-day gap). Because both groups were tested on Day 22, the spaced practice group had a shorter test delay than the massed practice group, a confounding that favored the former group.

Three experiments have assessed the spacing of mathematics practice. In a study by Rea and Modigliani (1985), third graders saw multiple presentations of multiplication facts that were spaced or massed. When the students were tested immediately afterward, spacing increased their test scores by 105%. Learning multiplication facts, however, requires only the verbatim memorization of paired associates. How does spacing affect practice with more challenging tasks?

In two recent spacing experiments, longer test delays were used, and the students learned a mathematical procedure. In the first of these studies (Rohrer & Taylor, 2007), college students observed a tutorial on the permutation task described above and then were randomly assigned to work practice problems that were either massed or spaced across two sessions separated by 1 week. Each practice problem was allotted a fixed amount of time and was followed immediately by a visual presentation of the complete solution. As assessed by a test given 1 week later, the spacing of practice problems boosted test scores (74% vs. 49%, Cohen's $d = .66$). In a similar experiment by the same authors (Rohrer & Taylor, 2006), a spacing effect was observed after a 4-week test delay (64% vs. 32%, Cohen's $d = .86$).

Spacing works by reducing the rate of forgetting, as evidenced by the fact that its benefits often increase with longer test delays (e.g., Dempster, 1988; Willingham, 2002). The effects of forgetting are often neglected by learning theorists, but acquisition has little utility unless material is retained. Indeed, although poor performance on standardized achievement tests is often attributed to the absence of acquisition, forgetting may often be the culprit. For example, in the 1996 National Assessment of Educational Progress, 50% of U.S. eighth graders were unable to correctly multiply −5 and −7, even though the question was presented in a multiple-choice format (Reese, Miller, Mazzeo, & Dossey, 1997). If any of these erring students knew the product previously, which seems likely, their error was likely due to forgetting.

In summary, the benefits of spacing have been observed over a wide range of ages, tasks, settings, and time periods, and particularly large effects have been found with mathematics tasks. Only three spacing experiments, however, have used mathematics tasks, and the generality of the mathematics spacing effect is therefore still unknown (as discussed further below). Still, in light of the hundreds of studies demonstrating the robustness of the spacing effect, and given the uniformly large benefits of spaced mathematics practice described above, it would seem that spaced practice is being grossly underutilized in mathematics instruction, as other authors have concluded (e.g., Bahrick & Hall, 1991; Dempster, 1988; Mayfield & Chase, 2002; Willingham, 2002).

Effect of Mixing

Whereas spacing provides temporal separation of problems on a single topic, mixed practice includes a variety of problems on different topics. Studies of both mathematical and nonmathematical tasks have found that mixing typically improves subsequent performance. As with spacing, the bulk of the mixing studies have relied on nonmathematical tasks (for a review, see Bjork, 1994). For instance, Kornell and Bjork (2008) taught students to recognize the styles of various artists by showing college students paintings by each artist. When the paintings by different artists were mixed rather than being blocked by artist, the students were better able to identify the artist in a subsequent test—even though the test consisted solely of paintings that the students had not seen. Thus, mixing improved the students' ability to discriminate styles.

This improved ability to discriminate suggests that mixing problems on different topics might be especially suited to learning mathematics because students often struggle to pair a mathematics problem with the appropriate procedure or concept. Whereas blocked practice often ensures that students will know the appropriate strategy for a problem before they read it, mixed review requires that they identify the appropriate strategy on the basis of the problem itself. Consequently, mixing improves discrimination ability because it gives students the opportunity to recognize which features of a problem are relevant to the choice of concept or procedure (e.g., Sweller, van Merrienboer, & Paas, 1998). For example, if a statistics lesson on the one sample *t* test is followed by a practice set composed solely of one-sample *t*-test problems, the appropriate choice of statistical test for each problem is obvious in advance, thereby allowing students to successfully complete the practice set without knowing why a particular test is appropriate. In other words, whereas blocked practice requires students to know *how* to perform a procedure, it does not require them to know *which* procedure is appropriate. This weakness is ultimately exposed, of course, when students must solve problems without the crutch that blocked practice provides, which is required, for instance, during midterm or final examinations and standardized tests.

Discrimination ability is difficult to acquire in mathematics, in part because superficially similar problems often require different strategies. For instance, the two integration problems

$$\int ex^e dx = ? \text{ and } \int xe^x dx = ?$$

resemble one another, but only the latter requires integration by parts. Moreover, the difficulty of identifying the appropriate concept or procedure is not limited to procedural problems. For example, the difficulty of a word problem is typically due in large part to the absence of an explicit reference to the appropriate concept or procedure, as in the following problem, which is similar to those in many textbooks:

> Two strings, one 28 inches long and another 70 inches long, are to be cut into shorter sections so that all sections (from both strings) are equal in length. What is the greatest possible length of each section?

This problem requires students to infer that they must find the greatest common factor of 28 and 70, whereas no such inference is needed for the purely procedural problem “What is the greatest common factor of 28 and 70?” The word problem loses most of its pedagogical value, however, if it appears in a set of problems on the greatest common factor because students can then correctly assume that it merely requires them to find the greatest common factor of the two integers given in the problem statement. With blocked practice, therefore, students can solve a word problem without reading any words.

The importance of discrimination ability was demonstrated in a set of experiments by VanderStoep and Seifert (1993). In one of their studies, for instance, college students were taught to solve two kinds of mathematics problems that were either superficially similar or dissimilar, and the instruction emphasized one of two skills:

learning *how* to solve each kind of problem or learning *which* of two procedures was appropriate. When the subsequent test included two superficially similar kinds of problems, students performed better if their instruction had emphasized "learning *which*" instead of "learning *how*." But when the two kinds of test problems were superficially dissimilar, the two instructional techniques produced roughly equal test scores.

Finally, a direct comparison of mixed and unmixed practice was the aim of two recent experiments. In the first (Rohrer & Taylor, 2007), college students learned how to find the volume of four obscure solids. Every student saw the same lessons and the same practice problems, and immediately after each practice problem, every student saw the correct solution. But by random assignment, the four kinds of practice problems (*a, b, c,* and *d*) were arranged in one of two ways:

Unmixed practice:	*a a a a*	*b b b b*	*c c c c*	*d d d d*
Mixed practice:	*a b c d*	*b d a c*	*c a d b*	*d c b a*

One week later, the students returned for a second practice session in which the four kinds of problems were arranged as in the first session, thereby ensuring that the practice of each problem kind was spaced across sessions. Scores on a test given 1 week later revealed that mixed practice tripled the test scores (63% vs. 20%, Cohen's $d = 1.29$).

In a similar experiment by Taylor and Rohrer (2008), 9-year-old and 10-year-old students learned and practiced four kinds of problems on prisms (e.g., find the number of edges of a prism given the number of sides of its base). Again, subsequent test performance was far greater after mixed practice than after unmixed practice (78% vs. 38%, Cohen's $d = 1.21$). Also, this study included a second test in which the appropriate formula for each problem was provided. Both groups performed nearly perfectly on the second test, which is consistent with the view that the benefit of mixing reflects an improved ability to pair problems and strategies.

Caveats and Limitations of the Experimental Data

Although the mathematics learning studies summarized above favor mixed review over blocked practice, those studies relied on a relatively narrow slice of procedures and tasks. Most notably, all of the studies employed tasks that, by one commonly drawn distinction (e.g., Rittle-Johnson & Alibali, 1999), were *procedural* (concerning the steps needed to solve problems), rather than *conceptual* (concerning the underlying principles). Or with respect to an alternative distinction put forth by Hatano and Inagaki (1986), the tasks required *routine expertise*, which allows students to solve problems "quickly and accurately without understanding," rather than *adaptive expertise*, which is an "ability to apply meaningfully learned procedures flexibly and creatively" (Hatano, 2003, p. xi). Notably, adaptive expertise requires both procedural and conceptual knowledge (e.g., Baroody, 2003). Only further research can determine whether mixed review engenders such expertise to a lesser or greater degree than blocked practice does.

A second limitation of the extant data is that mixed review and blocked practice have not been experimentally assessed in classroom settings. Several year-long classroom-based studies compared a Saxon textbook (e.g., Saxon, 1997), which relies primarily on mixed review, with other textbooks relying primarily on blocked practice. Although some of those studies found support for the Saxon textbooks (What Works Clearinghouse, 2007), none specifically assessed the effects of mixed review. Any two textbooks differ in numerous ways, ensuring that the result of any study comparing them cannot be logically attributed to just one of the differences between them (such as the arrangement of practice problems).

Finally, it is worth emphasizing that the choice between mixed review and blocked practice is a false dichotomy. In fact, a hybrid approach might be optimal. For instance, immediately after students encounter new material, blocked practice allows them to achieve some mastery of procedural skill before they encounter more complex problems that require them to apply the procedure in novel ways or make connections between the new material and previously learned material. Also, a blocked practice set can include problems that "build upon each other" in a manner intended to foster guided self-discovery, as exemplified by textbooks in the Connected Mathematics series (e.g., Lappan, Fey, Fitzgerald, Friel, & Phillips, 2004). Hence, the most advantageous arrangement of practice problems likely requires both mixed review and blocked practice.

RETRIEVAL PRACTICE

Mixed review is a blend of spacing and mixture that also increases students' reliance on a strategy known as *retrieval practice* in which the to-be-learned information must be retrieved from memory. With Spanish-English flash cards, for example, students attempt retrieval (e.g., *casa–?*) before seeing the answer (*house*), whereas no retrieval occurs during mere rereading (*casa–house*). Numerous studies have found benefits of retrieval practice with a variety of nonmathematical tasks, and, in fact, retrieval practice boosts retention even after retrieval fails, assuming that students see the correct answer soon after failure (for a review, see Roediger & Karpicke, 2006).

Mixed review increases the use of retrieval practice in two ways. First, and as discussed above students must retrieve the appropriate concept or procedure from memory. For example, upon seeing a quadratic equation that cannot be solved by factoring using integral coefficients, students must recall that the quadratic formula is necessary. Second, once the correct concept or procedure is recalled, mixed review requires that students recall how to apply the concept or procedure. For example, in using the quadratic formula, what do a, b, and c represent? In contrast, when students confront a series of problems involving the same concept or procedure, such retrievals are required for only the first of the problems.

The retrieval practice effect is, at first glance, at odds with studies showing that solving problems is sometimes less effective than merely reading worked examples (e.g., Sweller & Cooper, 1985). As an explanation, Sweller and his colleagues (e.g., Sweller et al., 1998) have argued that the reading of worked examples frees students from those demands of conventional problem solving that are extrinsic to the acquisition of the most critical information. (This is an application of cognitive load theory, which is discussed below.) Still, as Sweller et al. note, a fundamental disadvantage of reading worked examples is that students need not attend deeply to the material, and, to address that concern, these authors suggest the use of so-called completion problems, which require students to solve only one part of the problem (e.g., van Merrienboer, 1990). Thus, completion problems can ensure an efficient use of students' efforts while also providing the benefits of retrieval practice.

COGNITIVE DEMANDS OF MIXED REVIEW

The difference between blocked practice and mixed review is very salient to students, of course. Indeed, Paas and van Merrienboer (1994) found that mixed practice increased students' judgments of problem difficulty and the time they devoted to each practice problem. The increased difficulty can be assessed both empirically and theoretically.

Impaired Practice Performance

Although spacing and mixing often boost test scores, each feature impairs *practice performance*. For instance, in the mixed practice study by Rohrer and Taylor (2007) described above, the large test benefit of mixing (63% vs. 20%) occurred even though the mixed practice performance was worse than the unmixed practice performance (60% vs. 89%). A feature that decreases practice performance while increasing test performance has been described by Bjork and his colleagues as a *desirable difficulty*, and spacing and mixing are two of the most robust ones (e.g., Kornell & Bjork, 2008; for reviews, see Bjork, 1994; Schmidt & Bjork, 1992). As these researchers have noted, students and teachers sometimes avoid desirable difficulties such as spacing and mixing because they falsely believe that features yielding inferior practice performance must also yield inferior learning.

Cognitive Load

One theoretical analysis of the demands of mixed review is offered by *cognitive load theory* (for reviews, see Sweller, 1994; Sweller et al., 1998), which holds that the cognitive demands on a student reflect competition for *working memory*, which is the mental mechanism by which people temporarily store and manipulate information needed for complex cognitive tasks (Baddeley, 1992). Whereas a virtually unlimited amount of information can be stored in long-term memory, the capacity of working memory is very limited (e.g., Baddeley, 1992; Sweller et al., 1998). Consequently, working memory can be easily overwhelmed, and cognitive load theory posits that learning is optimized when working memory is used as efficiently as possible.

Spacing likely increases cognitive load if only because it typically requires more memory retrieval than blocked practice. For example, if a dozen Pythagorean theorem problems appear sporadically across several months rather than in immediate succession, each problem will require students to retrieve both the formula and how it should be used (e.g., replace *c* with the length of the hypotenuse). This retrieval from long-term memory requires working memory, as evidenced by many findings showing that performance on working memory tasks is impeded by a concurrent memory retrieval task (e.g., Baddeley, Lewis, Eldridge, & Thomson, 1984; Rohrer & Pashler, 2003). With practice, though, information from long-term memory is retrieved with greater ease, leading to a concomitant drop in cognitive load (e.g., Sweller et al., 1998).

The effect of mixed practice on cognitive load is more subtle. Sweller et al. (1998) state that variability in practice—which occurs with mixing—increases cognitive load, yet these authors also note that variability improves learning. Thus, although cognitive load theory typically explains better learning through the reduction of cognitive load, variability is said to increase cognitive load even as it improves learning. Sweller et al. (1998) refer to this outcome as a paradox, and they resolve it by distinguishing between different kinds of load. If learning is enhanced, the additional load is *germane*; if not, the additional load is *extrinsic*. (A third kind of load, *intrinsic load*, reflects the difficulty of the problem itself.) Thus, this aspect of the theory is definitional and not testable, because the effect of any manipulation on subsequent test performance can be explained post hoc. Still, the account elegantly explains certain findings. For instance, Paas and van Merrienboer (1994) found that variability across successive problems proved useful only when the problems were not too difficult. Thus, according to cognitive load theory, the variability was useful (i.e., germane).

The cognitive demands of mixed practice also can be assessed through its demands on students' attention. When people must attend to multiple features of the same task or perform different mental tasks at the same time, a considerable amount of research suggests that the resulting impaired performance reflects not the increased total demand but rather an inability to perform two tasks at once (for a review, see Pashler, 1994). Consequently, a *bottleneck* arises, and people must either switch back and forth or finish one task before beginning the other.

This alternation between different kinds of tasks typically produces a *task-switching cost* (for a recent review, see Monsell, 2003). For instance, in a series of experiments by Rubinstein, Meyer, and Evans (2001), learners repeatedly performed two kinds of tasks, such as arithmetic computation or the classification of geometric shapes. Compared with blocked practice, alternation (*a, b, a, b,* etc.) dramatically slowed performance, and this task-switching cost increased with task complexity. Whereas solving a dozen successive problems of the same kind allows students to acquire a sort of cognitive momentum, the repeated shifting of attention required by mixing might also contribute to the cognitive demand of mixed practice.

MIXED REVIEW AND THE TEACHER

Naturally, the use of mixed review also alters the demands placed on the teacher. Most notably, perhaps, teachers who devote a portion of the lesson to a discussion of a set of yesterday's homework problems that provided mixed review will likely encounter a greater number and variety of questions from students than if the set had provided only blocked practice. On the other hand, mixed review ensures that students who fail to understand a lesson (or fail to attend the class) will still encounter many problems on topics they have seen previously, whereas blocked practice ensures that such students will have less success. Likewise, mixed review provides students with more time to grasp newly introduced material, allowing them to seek help in a following lesson, if necessary, before seeing more challenging problems in subsequent practice sets.

The use of mixed review also requires that teachers who omit a single topic must be sure to omit every problem relying on that topic, which is not a trivial task when those problems are distributed across many practice sets. For that reason, textbooks using mixed review typically preface each practice problem with a brief reference to the corresponding lesson number, which enables students and instructors to quickly find the relevant lesson. Alternatively, listing those references elsewhere in the textbook would reduce the chance that the student will either consult the appropriate lesson before attempting the problem or learn to recognize a problem on the basis of its lesson number alone.

It should be emphasized, however, that the use of mixed review does not require any changes in how the teacher presents new material. That is, although mixed review ensures that practice sets will span a variety of topics, the introduction of each lesson can be allotted the usual amount of time, allowing for a thorough presentation and numerous practice problems in class. An increased reliance on mixed review is likely to have less of an effect on the instructor than on the student.

CONCLUSION

Experiments have shown that test scores can be dramatically improved by the introduction of spaced practice or mixed practice, which are the two defining features of mixed review. Moreover, neither spacing nor mixing requires an increase in the number of practice problems, meaning that both features increase efficiency as well as effectiveness. These effects, however, have been assessed in a relatively small number of studies, and it remains unknown how far they generalize. Still, in light of the uniformly large effects of spacing and mixing that have been observed, and in light of the ease with which mixed review can be incorporated into a textbook or computer-assisted instructional program, its effects on mathematics learning deserve greater consideration by teachers and researchers.

REFERENCES

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., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120*, 20–33.

Baddeley, A. (1992). Working memory. *Science, 255*, 556–559.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113,* 518–540.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–34). Mahwah, NJ: Erlbaum.

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

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*, 245–248.

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. (in press). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*.

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.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615–622.

Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology, 59,* 290–296.

Fitts, P. M. (1965). Factors in complex skill training. In R. Glaser (Ed.), *Training research and education* (pp. 177–197). New York: Wiley.

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

Glencoe. (2001). *Mathematics: Applications and connections–Course 1*. New York: Author.

Hall, F. H. (1989). *Learning and memory.* Boston: Allyn & Bacon.

Hatano, G. (2003). Foreword. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. xi–xii). Mahwah, NJ: Erlbaum.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York: Freeman.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*, 71–78.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (2004). *Connected mathematics*. Needham, MA: Pearson.

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

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

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134–140.

Paas, F., & van Merrienboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning: Journal of Practical Research & Applications, 4*, 11–18.

Reese, C. M., Miller, K. E., Mazzeo, J., & Dossey. J. A. (1997). *NAEP 1996 Mathematics Report Card for the Nation and the States.* Washington, DC: National Center for Education Statistics.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55,* 297–308.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*, 175–189.

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

Rohrer, D., & Pashler, H. (2003). Concurrent task effects on memory retrieval. *Psychonomic Bulletin & Review, 10*, 96–103.

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.

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 763–797.

Saxon, J. H., Jr. (1997). *Algebra I* (3rd ed.). Norman, OK: Saxon Publishers.

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.

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

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295–312.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59–89.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.

Taylor, K., & Rohrer, D. (2008). *The effects of interleaved practice.* Unpublished manuscript, University of South Florida.

Toppino, T. C., Kasserman, J. E., & Mracek, W. A. (1991). The effect of spacing repetitions on the recognition memory of young children and adults. *Journal of Experimental Child Psychology, 51*, 123–138.

University of Chicago School Mathematics Project. (2007). *Everyday math* (3rd ed.). Chicago: Wright Group/McGraw-Hill.

VanderStoep, S. W., & Seifert, C. M. (1993). Learning "how" versus learning "when": Improving transfer of problem-solving principles. *Journal of the Learning Sciences, 3,* 93–111.

Van Merrienboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6,* 265–287.

What Works Clearinghouse. (2007, April). *Saxon middle school math* (WWC Intervention Report, Middle School Math). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved August 18, 2008, from http://ies.ed.gov/ncee/wwc/pdf/WWC_Saxon_Math_Middle_040907.pdf

Willingham, D. T. (2002, Summer). Allocating student study time: Massed vs. distributed practice. *American Educator*, 37–39, 47.

Author

Doug Rohrer, Department of Psychology, PCD 4118G, University of South Florida, Tampa, FL 33620; drohrer@cas.usf.edu