

Interleaved Practice Improves Mathematics Learning

Doug Rohrer, Robert F. Dedrick, and Sandra Stershic
University of South Florida

A typical mathematics assignment consists primarily of practice problems requiring the strategy introduced in the immediately preceding lesson (e.g., a dozen problems that are solved by using the Pythagorean theorem). This means that students know which strategy is needed to solve each problem before they read the problem. In an alternative approach known as *interleaved practice*, problems from the course are rearranged so that a portion of each assignment includes different kinds of problems in an interleaved order. Interleaved practice requires students to choose a strategy on the basis of the problem itself, as they must do when they encounter a problem during a comprehensive examination or subsequent course. In the experiment reported here, 126 seventh-grade students received the same practice problems over a 3-month period, but the problems were arranged so that skills were learned by interleaved practice or by the usual blocked approach. The practice phase concluded with a review session, followed 1 or 30 days later by an unannounced test. Compared with blocked practice, interleaved practice produced higher scores on both the immediate and delayed tests (Cohen's $d_s = 0.42$ and 0.79 , respectively).

Keywords: learning, mathematics, interleaved, spaced, practice

The ongoing effort to improve students' mathematics proficiency has included a wide variety of interventions. Some are rooted in theoretical models of cognition, some are designed to eliminate misconceptions, and still others aim to reduce students' mathematics anxiety. Here we describe an intervention that is inspired by what is perhaps the simplest principle of learning: the practice of a skill improves the performance of that skill. It might seem that this robust maxim is already widely used in the mathematics classroom because nearly all teachers assign practice problems like the ones appearing on tests. However, most practice assignments are arranged in a way that simplifies the solution to each problem, and this crutch is usually not available to students when they are tested.

To see why this is the case, we must first point out that the solution to nearly any mathematics problems includes two distinct steps, as illustrated by the following problem.

A girl hikes 8 km east and then 15 km north. How far is she from her starting point?

This problem is solved by the Pythagorean theorem (the answer is 17 km because the unknown distance is the hypotenuse of a right triangle, and $8^2 + 15^2 = 17^2$). However, before students can use

the Pythagorean theorem to solve this problem, they must first infer that they *should* use it. In other words, the solution to a mathematics problem requires students to choose a strategy, not only execute the strategy, and students often find the choice of strategy to be more challenging than its execution (e.g., Kester, Kirschner, & van Merriënboer, 2004; Siegler, 2003; Siegler & Shrager, 1984). In formal terms, the choice of an appropriate strategy requires students to both discriminate between different kinds of problems (i.e., problems requiring different strategies) and associate each kind of problem with an appropriate strategy (Figure 1).

The choice of an appropriate strategy is often difficult because superficially similar problems sometimes require different strategies (e.g., Chi, Feltovich, & Glaser, 1981; Siegler, 2003). For example, word problems often lack explicit cues that indicate the kind of problem it is. For example, the previous word problem required the Pythagorean theorem, but the problem did not include any mention of the Pythagorean theorem or the words *triangle* or *hypotenuse*. In algebra, students must solve equations, but the instruction "Solve for x " does not indicate which of the different strategies is useful (e.g., factoring, quadratic formula, and so on). Similarly, much of calculus is devoted to integration, and students must learn to discriminate between problems that look alike yet require different integration techniques (e.g., $\int e^x dx$ and $\int x e^x dx$). In short, students at nearly every level of mathematics must learn to discriminate between different kinds of problems with similar surface features.

Yet students need not learn to choose a strategy when every problem within a practice assignment requires the same strategy—an approach known as *blocked practice*. With blocked practice, students know the strategy before they read the problem. For example, if a lesson on proportions is followed by a dozen word problems requiring students to create a proportion, students need not learn which features of a problem indicate that it can be solved by a proportion. In short, blocked practice does not provide stu-

Doug Rohrer, Department of Psychology, University of South Florida; Robert F. Dedrick, Department of Educational and Psychological Studies, University of South Florida; Sandra Stershic, Department of Psychology, University of South Florida.

This work was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110517 (principal investigator: Doug Rohrer). The opinions expressed are those of the authors and do not necessarily represent the views of the U.S. Department of Education. We thank Hillsborough County Public Schools for their participation.

Correspondence concerning this article should be addressed to Doug Rohrer, Psychology PCD4118G, University of South Florida, Tampa, FL 33620. E-mail: drohrer@usf.edu

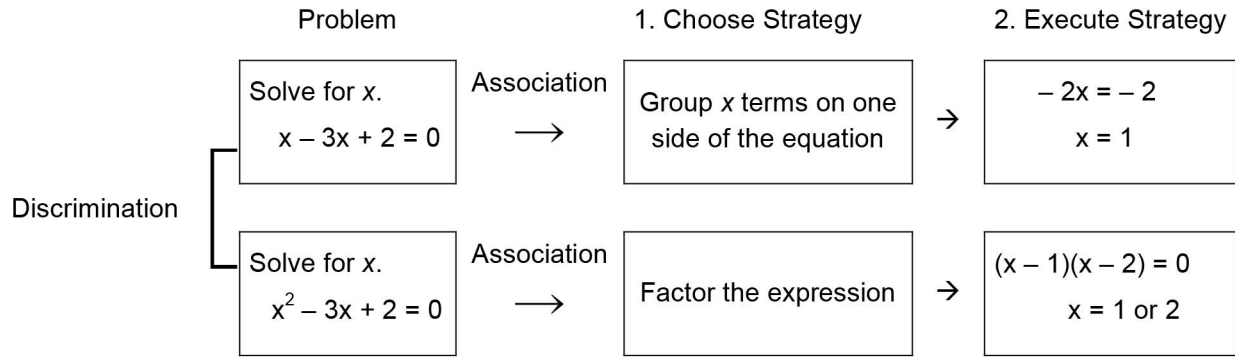


Figure 1. The solution of a mathematics problem. Students must learn to choose a strategy and execute the strategy. The choice of an appropriate strategy requires that students discriminate between different kinds of problems and associate each kind of problem with an appropriate strategy. Students do not need to choose a strategy if every practice problem within an assignment requires the same strategy (*blocked practice*).

dents with an opportunity to choose an appropriate strategy on the basis of the problem itself, and yet they must perform this skill when they sit for an exam consisting of multiple kinds of problems.

Blocked practice can dramatically reduce the difficulty of a mathematics problem, partly because students need not discriminate between problems requiring different strategies. Consider this example.

Chloe baked 24 cookies. Her father ate 10 of the cookies. How many cookies remain?

This problem is solved by simple subtraction ($24 - 10 = 14$), but students who are capable of subtraction may not be able to infer that they *should* subtract, possibly because the problem does not include cues such as the words *subtract* or *difference*. However, if this problem follows other problems requiring subtraction, students know the strategy in advance. In fact, they could solve this problem without reading anything other than the two integers (24 and 10). Put another way, blocked practice sometimes allows students to solve word problems without reading any words.

Aside from excusing students from having to discriminate between different kinds of problems, blocked practice can also impede the learning of the association between a problem and an appropriate strategy. For example, the instruction “Find the perimeter” indicates unambiguously that the problem is solved by adding the lengths of the sides of the given polygon (e.g., a square with sides of length 3), and this salient feature of the problem presumably makes it easier for students to recognize what kind of problem it is (perimeter problem). Yet, if a practice assignment includes a block of perimeter problems, students can solve each problem without reading or attending to the instruction “Find the perimeter.” This repeated failure to perceive the word *perimeter* weakens the association between the kind of problem (perimeter) and the strategy (add the lengths of the sides). In effect, blocked practice allows students to complete an assignment without being aware of the kind of problem they were solving.

Blocked mathematics practice is prevalent. We inspected several commonly used middle school mathematics textbook series and found that most of the practice problems in each textbook appear in a block of problems devoted to the same concept or procedure. In particular, the prototypical assignment may in-

clude problems in a variety of formats (e.g., procedural problems, word problems, open-ended questions, and standardized test practice problems), but most of the practice problems are dedicated to the immediately preceding lesson. In some of the textbooks, each practice assignment consists of many dozens of problems (sometimes more than 100, presumably so that teachers can choose a subset of problems), and the latter portion of each assignment includes a group of between five and 10 problems drawn from previous lessons and grouped within a section labeled *Mixed Review* or *Spiral Review*. However, these review problems comprised less than 15% of the total number of practice problems in our informal survey. It is also true that many textbooks include periodic review assignments, usually at the end of a chapter, although these assignments typically include a small block of problems for each lesson (e.g., a few problems based on the first lesson of the chapter, followed by a few problems based on the second lesson, and so forth). Moreover, the traditional mathematics textbook is increasingly supplemented or replaced by a “consumable workbook,” in which students are asked to write their solutions on tear-away pages, and our survey indicates that these workbooks rely more heavily on blocked practice than do the textbooks. Unfortunately, the predominance of blocked practice cannot be precisely gauged because teachers may omit parts of the materials adopted by their school or school district, or they might choose their own materials (e.g., assignments freely downloaded from the Internet). Nevertheless, it appears that the vast majority of mathematics students devote most of their practice effort to blocked practice. For that reason, blocked practice served as the control in the present study.

In an alternative approach that served as the intervention in the present study, practice problems are merely rearranged so that a portion of each assignment includes a set of different kinds of problems presented in an intermixed order—a technique known as *interleaved mathematics practice* (e.g., Higgins & Ross, 2011; Richland, Bjork, Finley, & Linn, 2005; Richland, Linn, & Bjork, 2007; Rohrer & Taylor, 2007; Schmidt & Bjork, 1992). With interleaved practice, students must learn to choose the strategy on the basis of the problem itself. One hypothetical illustration of interleaved mathematics practice is shown in Figure 2.

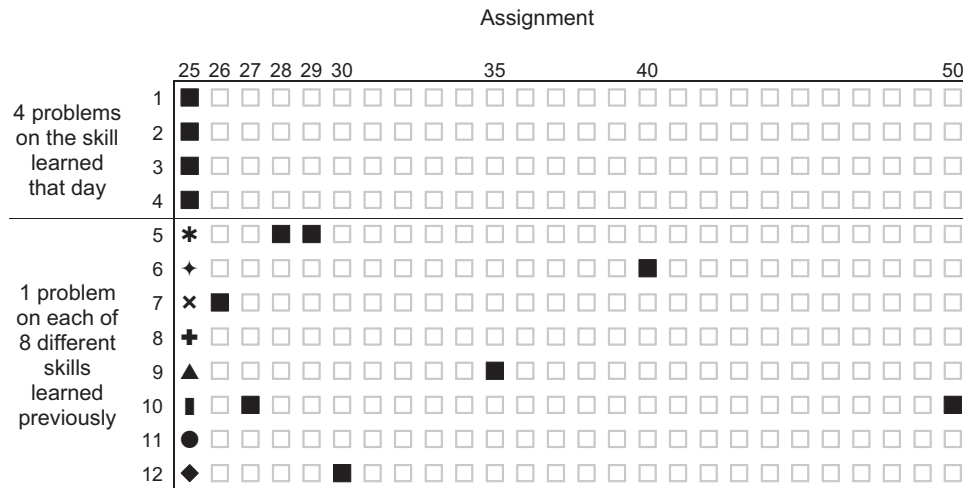


Figure 2. Interleaved mathematics practice. In this hypothetical illustration, Assignment 25 includes a small block of four problems (squares) and eight problems of interleaved practice (non-squares). For example, if Assignment 25 follows a lesson on proportions, it would include four problems on proportions and one problem on each of eight different skills learned earlier in the course or during a prior course. Eight additional proportion problems are distributed across subsequent assignments. The empty squares are placeholders for unidentified kinds of problems.

In addition to the juxtaposition of different kinds of problems within the same assignment, interleaved practice guarantees that problems of the same kind are distributed or *spaced* across different assignments (Figure 2). Dozens of studies have demonstrated that spacing improves classroom learning, and the effect of spacing is one of the largest and most robust effects in the learning literature (e.g., for recent reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Küpper-Tetzel, 2014; Roediger & Pyc, 2012). Furthermore, several studies have shown that spacing can improve the learning of mathematics (Bahrck & Hall, 1991; Gay, 1973; Grote, 1995; Rohrer & Taylor, 2006, 2007; Yazdani & Zebrowski, 2006). In the creation of interleaved mathematics assignments, the spacing intervals between consecutive problems of the same kind might expand, meaning that a particular kind of problem might appear initially in each of several consecutive assignments and then turn up with decreasing frequency thereafter, as in the hypothetical illustration in Figure 2. Alternatively, the spacing interval between consecutive problems of the same kind may be fixed so that problems of the same kind appear once every, say, week or two, as in the present study. (There is some debate about whether equal or expanding spacing intervals are more effective, and the optimal choice seemingly depends on a variety of circumstances, e.g., Balota, Duchek, & Logan, 2007; Küpper-Tetzel, 2014; Storm, Bjork, & Storm, 2010). To summarize, interleaved mathematics practice has two beneficial features: problems of different kinds are juxtaposed, which requires students to choose a strategy, and problems of the same kind are spaced, which improves retention.

The earliest studies of interleaved practice examined its effects on skill learning (e.g., Hall, Domingues, & Cavazos, 1994; Shea & Morgan, 1979). For instance, in the study reported by Hall et al., college baseball players practiced hitting three types of pitches (fastball, curveball, and change-up) that were either blocked by type or interleaved, and interleaved practice led to better hitting on

a final test requiring batters to hit pitches of all three types without knowing the type of pitch in advance, just as they would need to do in a game. This kind of batting test is analogous to a mathematics test in which students do not know the kind of problem in advance.

With regard to mathematics learning, we know of five studies of interleaved practice, and four of these studies were conducted in a nonclassroom setting. In the first of these studies, Mayfield and Chase (2002) had remedial college students learn several simple algebraic rules (e.g., $x^5 x^2 = x^7$) with a practice schedule that provided either interleaved or blocked practice, and interleaved practice produced superior test scores (effect size unknown). Although the two groups of students in this study did not receive exactly the same problems or even the same number of problems because the study was not designed to assess interleaved practice, we believe it is the first demonstration of a mathematics interleaving effect. In a later study that explicitly compared interleaved and blocked practice, Rohrer and Taylor (2007) taught college students to find the volume of several obscure solids (e.g., spheroid and spherical cone) and found a positive interleaving effect on a test given 1 week later (Cohen's $d = 1.34$). This finding was later replicated (with the same materials but a different procedure) by Le Blanc and Simon (2008), who found a large interleaving effect ($\eta_p^2 = .32$) and further observed that interleaving improved students' ability to predict their test scores. Finally, Taylor and Rohrer (2010) taught fourth-grade students how to solve problems relating to prisms (e.g., "Find the number of faces on a prism with a 5-sided base"), and they found that a session of interleaved practice (rather than blocked practice) led to greater scores on a test given 1 day later ($d = 1.21$).

Most recently, Rohrer, Dedrick, and Burgess (2014) assessed the effects of interleaved mathematics practice in a classroom-based experiment. Seventh-grade students received a dozen practice problems of each of several kinds that were interleaved or

blocked, and interleaved practice produced greater scores on a final test ($d = 1.05$). Unlike the kind of problems used in previous studies of interleaving, the different kinds of problems were superficially dissimilar (e.g., proportion word problems, solving a linear equation, graphing equations), showing that the benefits of interleaving are not necessarily limited to scenarios in which students see only similar kinds of problems like those chosen in the laboratory studies summarized previously.

The present study differed from previous studies of interleaved mathematics practice in two fundamental ways. First, in the present study, the last practice assignment was followed by a review. Though seemingly innocuous, the inclusion of a review addresses a weakness that is inherent in a design comparing interleaved and blocked practice. Without a review, the use of interleaved rather than blocked practice intrinsically shortens the delay between the last practice problem of each kind and the test, as illustrated by the hypothetical illustration in Figure 2. This confounding works in favor of interleaved practice because shorter test delays improve test scores. The inclusion of a review in the present study therefore equated the delay between the final practice problem and the test, as shown in Figure 3. However, even with a review, most of the practice problems appeared later in the experiment if practice problems were interleaved rather than blocked, although this difference is arguably an intrinsic benefit of interleaved practice.

Furthermore, the inclusion of a review in the present study means that the blocked practice condition is a more suitable counterfactual. This is because even teachers who rely heavily on blocked practice assignments often give their students a review before a cumulative exam or high-stakes test. For instance, although students might have added fractions in only the first week of the school year, a review provides one more dose shortly before

the test. This in turn guarantees that these students benefit from spaced practice. In brief, the inclusion of a review ensures that the counterfactual used in the present study will be more effective and more characteristic of prevailing teaching practices than that used in previous studies of interleaving.

The second difference between the present study and previous interleaving studies is that the present one included an independent manipulation of test delay (1 or 30 days). This was done so that we could assess whether the benefits of interleaved practice decrease over time. Such a finding would dramatically curtail the utility of interleaved practice, and this possibility is not a straw man hypothesis. In fact, even some recommended learning strategies produce a boost in test scores that decreases markedly within a few weeks (e.g., Driskell, Willis, & Cooper, 1992).

In the experiment reported here, seventh-grade students saw their teachers' usual lessons and received assignments that we created. Every student received the same problems, but the scheduling of the problems was altered so that students received blocked or interleaved practice. Later, students received a review, followed 1 or 30 days later by an unannounced test.

Method

Participants

The study took place at a large public middle school in Tampa, Florida, during the 2013–2014 school year. Three mathematics teachers and nine of their seventh-grade classes participated. Each of the teachers had taught middle school mathematics for more than 5 years. The participating classes included only students who had received a passing score (3 or higher on a scale from 1 to 5) on the Grade 6 mathematics section of the 2013 Florida Comprehensive Assessment Test (FCAT, Version 2.0; Florida Department of Education, Bureau of K–12 Assessment, 2014), which they had taken in April 2013, near the end of the previous school year. A passing score on this test was achieved by 58% of the sixth-grade students at the participating school and by 52% of the sixth grade students in the state (Florida Department of Education, 2013).

Participation in the study required documentation of parent permission and student assent, which we received from 150 of the 164 students in the participating classes. Of these 150 students, 126 (84%) attended mathematics class on both the day of the unannounced review and the day of the unannounced test, and only their data were analyzed. Thus, the final sample included 126 students. Nearly every student was 12 years old at the beginning of the study, and 61 were girls (48%).

We did not ask students for information about their socioeconomic background or ethnicity because of privacy concerns, but the school district provided demographic data for the sample in aggregate. By these data, 38% of the students received free or reduced-price lunch, and the sample was ethnically diverse (10% Asian, 14% Black, 22% Hispanic, 6% multiracial, and 47% White).

Materials

Students received graph problems and slope problems, and no student saw the same problem more than once during the experiment. Graph problems required students to graph a linear equation

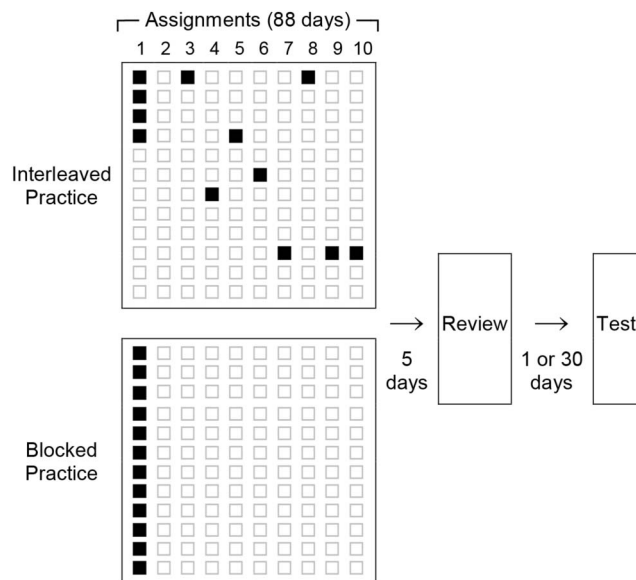


Figure 3. Procedure: The 10 assignments included 12 graph problems and 12 slope problems. The 12 problems of each kind were either grouped into a single assignment (blocked practice) or distributed across multiple assignments (interleaved practice). The dark squares indicate the location of the graph problems. The location of the slope problems is given in Appendix A.

of the form, $y = mx + b$, where m and b were nonzero single-digit integers. Examples include $y = 2x - 1$, $y = -x + 3$, and $y = -3x + 2$. Each problem included the instruction “Graph the equation” and was accompanied by a Cartesian grid. Students were permitted to use any appropriate method, but they were taught to find points by substituting at least two x values into the equation and find the corresponding y values. For example, for the equation $y = 2x - 1$, the substitution of $x = 2$ yields $y = 3$. For slope problems, students found the slope of the line passing through two given points on the Cartesian plane. Each problem began with the instruction, “Find the slope of the line that passes through the points” followed by a pair of points such as “(1, 5) and (8, 9)” or “(3, 1) and (9, -4).” Students were taught to find the slope by calculating $\Delta y/\Delta x$, which is known colloquially as “rise over run.” For example, the line passing through points (1, 5) and (8, 9) has slope $4/7$. In every slope problem, the two given points had integer coordinates, and the slope equaled a nonzero fraction between -1 and 1 .

Design

We manipulated practice schedule (interleaved or blocked) and test delay (1 or 30 days). Test delay was manipulated by randomly assigning each student within each class to either the 1-day or 30-day delay ($n = 63$ for each group). This meant that each class included students at both test delays.

Practice schedule was a counterbalanced within-subject variable. Students in Group 1 received interleaved practice of graph problems and blocked practice of slope problems, and Group 2 received the reverse. Group 1 ($n = 59$) included four classes (two taught by Teacher A, and two taught by Teacher B). Group 2 ($n = 67$) included five classes (two by A, two by B, and one by C). Classes were assigned to groups as follows. Two of the classes were designated as “honors/gifted” by the school, and these two classes were randomly assigned to different groups. The remaining seven classes were deemed by the school as being at the same level, and each of these classes was randomly assigned to one of the two groups with the constraint that teachers with more than one participating class had an equal number of classes in each group. The two groups scored similarly well on a test consisting of six multiple-choice problems from the Grade-8 mathematics portion of the National Assessment of Educational Progress, or NAEP (National Center for Education Statistics, 2013), 76% ($SD = 22\%$) vs. 79% ($SD = 22\%$), $t(108) = 0.62$, $p = .54$, Cohen’s $d = 0.12$.

Procedure

The study consisted of 10 practice assignments, a review session, and a test. Each practice assignment consisted of 12 problems presented on two sides of a single sheet of paper. The 10 assignments included 12 graph problems and 12 slope problems, and the remaining problems were drawn from unrelated topics (fractions, proportions, percentages, statistics, and probability). Teachers presented a tutorial on the graph problems immediately before giving Assignment 1, which included the first four graph problems, and they presented a tutorial on the slope problems immediately before giving Assignment 2, which included the first four slope problems. However, the scheduling of the remaining eight graph and eight slope problems varied. With blocked practice, students saw the

remaining eight problems immediately, which is to say that all 12 problems of a particular kind (graph or slope) appeared in the same assignment. With interleaved practice, the remaining eight problems were distributed across subsequent assignments (Figure 3 and Appendix A).

Students received the 10 assignments on Days 1, 6, 14, 32–33, 33 or 35, 35 or 38, 45–46, 72–75, 81–82, and 86–88. Students were asked to complete each assignment before the following school day, and the final practice assignment was collected by teachers on Days 87–89. On the due date for each assignment, teachers presented the solution to every problem with the aid of a slide show created by the authors. As teachers presented the solutions, students were asked to correct their errors. Teachers then collected the assignments. Within three school days, one or more authors visited the school and scored each student’s assignment without making any marks on the assignments. The assignments were then returned to the teachers.

Students’ scores on the practice assignments do not provide a valid measure of learning because students corrected their solutions before giving their assignments to their teachers. Even if teachers had collected the assignments at the beginning of class, the students might have received help from their parents or other students. This ambiguity is typical of students’ mathematics assignments, and many teachers encourage students to seek help with practice assignments.

Yet the scoring of the practice assignments provided an objective measure of the fidelity of the intervention (which consisted solely of the assignments). Most important, these scoring visits to the school revealed that each teacher distributed each of the 10 assignments to their students, and students’ self-corrected solutions further demonstrated that the teachers presented the solutions to the practice assignments. (Of course, this perfect rate of teacher compliance might have been achieved *because* we collected the assignments.) The scoring of the assignments also provided a rough measure of student compliance. When the graph or slope problems were blocked, students averaged 81% correct. In the interleaved practice condition, students averaged 84% correct, and they averaged 82% correct for the last eight of the 12 problems (which were the only problems that were part of an interleaved assignment, as shown in Figure 3). Thus, by this measure, the intervention and the counterfactual produced nearly equal rates of student compliance.

On Days 66–69, every student received a large set of review problems created by the school district, and this review assignment was designed to prepare students for a standardized semester exam given to all students in the district. (The exam is taken on a computer without the presence of a teacher, and the teachers do not see the test items in advance.) The review assignment included two problems on the graphing of a line and four problems on slope, though the problems were stated differently than the ones in the study.

Before the completion of the practice phase, teachers completed an anonymous survey about their views on interleaved practice. Each of the three teachers received a paper copy of the survey on Day 66. We visited the school on Day 75 and collected an envelope with their completed surveys. The survey is shown in Appendix B.

Every student received the same review on Day 93, about 5 days after teachers collected the last practice assignment (Days 87–89,

depending on the class). The review included one graph problem, one slope problem, and eight unrelated problems. The graph and slope problems were the sixth and seventh problem, respectively. Students received the review assignment at the beginning of the class period, and teachers presented the solutions at the end of the same class period. Teachers then collected the assignment from students, and one of the authors collected the assignments from the teachers at the end of that school day. The mean scores on the review problems were slightly higher, but not statistically so, for problems learned by interleaved practice rather than by blocked practice (94% vs. 89%). Again, we present these data as a measure of compliance rather than a measure of learning because, as with the practice assignments, students were given an opportunity to correct their solutions before turning in their assignments.

Students were tested 1 or 30 days after the review. Students were tested during their regular class meeting, and the teacher and one author proctored the administration of each test. For each of the two test dates (1 and 30 days after the review), students who were not scheduled to receive the test received a filler test consisting of six math problems unrelated to the experiment, all of which were drawn from the Grade-8 mathematics section of a recent NAEP (National Center for Education Statistics, 2013). We asked teachers not to inform students of the tests in advance, and teachers did not see any test problems until the experiment was completed. The test booklet included a cover sheet, a sheet of paper with three graph problems, and a sheet of paper with three slope problems. We created six versions of the test by reordering the problems within each page, and the page with the slope problems preceded the page with the graph problems in three of the six versions. None of the problems had appeared in either a practice assignment or the review. Students were allotted 18 min to complete the test and allowed to use their school-supplied basic calculator. Each test was scored at the school on the day of the test by two raters who were blind to condition. The two raters independently scored each answer as correct or not and later resolved the few discrepancies (six in 756). Internal consistency reliabilities, as measured by Cronbach's alpha, were .89 and .77 for the slope and graphing problems, respectively.

Results

Interleaved practice produced higher test scores than did blocked practice (Figure 4). Students tested 1 day after the review showed a moderate benefit of interleaving, 80% ($SD = 33\%$) vs. 64% ($SD = 42\%$), $t(62) = 2.39$, $p = .02$, $d = 0.42$, 95% confidence interval (CI) [0.07, 0.77]. Students tested 30 days after the review showed a large benefit of interleaving, 74% ($SD = 39\%$) vs. 42% ($SD = 43\%$), $t(62) = 4.54$, $p < .001$, Cohen's $d = 0.79$, 95% CI [0.43, 1.15]. A two-way analysis of variance (with practice schedule as a within-subject variable and test delay as a between-subjects variable) showed that interleaved practice was superior to blocked practice, $F(1, 124) = 24.43$, $p < .001$, $\eta_p^2 = .165$, and that test scores were greater at the shorter test delay, $F(1, 124) = 7.69$, $p < .01$, $\eta_p^2 = .058$. However, the interaction between practice schedule and test delay was not statistically significant, $F(1, 124) = 2.84$, $p = .09$.

Secondarily, we analyzed the effect of interleaved practice on test scores for the graph and slope problems separately for both the 1-day and 30-day test delays, resulting in four different compari-

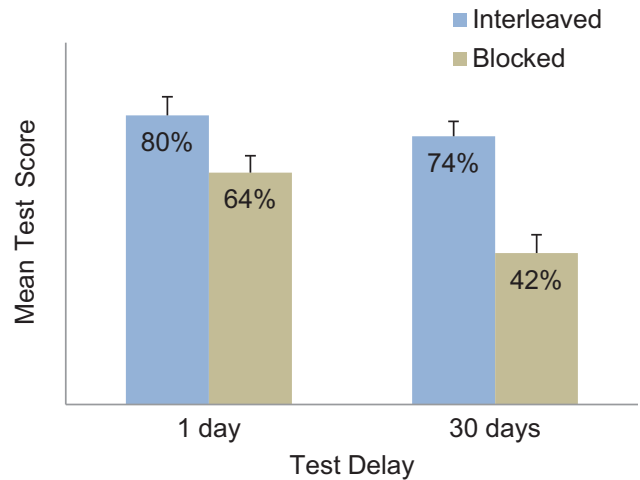


Figure 4. Results: Interleaving produced greater test scores at both test delays. Error bars represent 1 standard error. See the online article for a color version of this figure.

sons. Each difference was assessed by an independent t test, and in three of the four cases, the assumption of equal variances was rejected. We therefore did not use the pooled estimate for the error term for the t -statistic, and we adjusted the degrees of freedom using the Welch-Satterthwaite method. This correction did not affect the outcome of any of the four null hypothesis tests. For the 1-day test delay, interleaved rather than blocked practice resulted in higher mean test scores for the graph problems, 89% ($SD = 18\%$) vs. 73% ($SD = 36\%$), $t(50.6) = 2.25$, $p = .029$, $d = 0.54$, but the effect of interleaved practice on the slope problems was positive but not statistically significant, 73% ($SD = 41\%$) vs. 54% ($SD = 47\%$), $t(56.0) = 1.67$, $p = .10$, $d = 0.43$. For the test given after a 30-day delay, there was a statistically significant benefit of interleaving for both graph problems 84% ($SD = 30\%$) vs. 54% ($SD = 44\%$), $t(56.7) = 3.28$, $p = .002$, $d = 0.81$, and slope problems, 65% ($SD = 43\%$) vs. 29% ($SD = 39\%$), $t(61) = 3.44$, $p = .001$, $d = 0.87$.

The results of the teacher survey are shown in Appendix B. Although teachers responded anonymously and were encouraged to give their honest assessments, the sample size of only three teachers means that these data are speculative at best. Still, each teacher indicated that he or she would “use the intervention in the future if it was an option” and “would recommend the intervention to other math teachers.” All three also reported that interleaved practice did not “interfere with how [they] usually teach” and that “going over assignments” was no harder when assignments were interleaved rather than blocked. However, their opinions were split when they were asked whether their students found interleaved practice to be less likable, more challenging, or more time-consuming than blocked practice.

Discussion

The present study explicitly compared interleaved and blocked mathematics practice in a classroom setting and found that interleaved practice produced superior scores on a final test given 1 or 30 days later. Put another way, the mere rearrangement of practice

problems improved mathematics learning in the classroom. The study is also the first to demonstrate that the test benefit of interleaving does not diminish over time and perhaps grows larger. Finally, apart from its superiority to blocked practice, interleaved practice provided near immunity against forgetting, as the 30-fold increase in test delay reduced test scores by less than a tenth (from 80% to 74%).

Although the size of the effects might seem surprisingly large for a study conducted in a classroom, the effect sizes observed here are nevertheless much smaller than interleaving effects observed in the laboratory. Whereas the effects in the present study were medium to large ($d_s = 0.42$ and 0.79), laboratory studies of interleaving have uniformly found larger effects ($d = 1.34$, $d = 1.21$, and $\eta_p^2 = .32$; see introduction). In brief, the present findings are an instantiation—not a violation—of the adage that an intervention loses some of its efficacy when it is moved from the laboratory to the classroom (e.g., Hulleman & Cordray, 2009).

Another reason for the large effects of interleaving observed here and elsewhere is that interleaved mathematics practice inherently guarantees that students space their practice. That is, in addition to the juxtaposition of different kinds of problems within an assignment, problems of the same kind are spaced across assignments. However, the review session in the present study meant that even the blocked practice condition provided spacing, although to a lesser degree than that provided by the interleaved practice condition (Figure 3). In brief, the large effect observed here probably reflects the spacing effect, which is an inherent benefit of interleaved mathematics practice, but the contribution of spacing might have been reduced by the use of a review session.

The large effects notwithstanding, the present study has limitations. For instance, although the test problems were novel, the test problems and practice problems had the same format, and the observed effects might have been smaller if the test problems had required a greater degree of transfer. Also, the test benefit of interleaving might have been reduced if the review had included more than one problem of each kind (graph problem and slope problem), simply because a more intensive review session might have benefitted the blocked practice condition more than it did the interleaved practice condition. More broadly, it remains unknown whether the interleaving effects observed here would be found in a study with a wider variety of material and a greater number of teachers and students. Still, the ecological validity of the present study was reasonably good. Students learned from their teachers, the learning phase lasted 3 months, and the 1-month test delay was educationally meaningful.

We also should emphasize that the findings reported here do not suggest that blocked practice be avoided entirely. In fact, a small block of problems might be optimal, especially at the outset of an assignment given immediately after students are introduced to that kind of problem, perhaps because it gives students an opportunity to focus on the execution of a strategy (e.g., procedural steps and computation). Yet students who work more than a few problems of the same kind in immediate succession are likely to receive sharply diminishing returns on their additional effort (e.g., Rohrer & Taylor, 2006; Son & Sethi, 2006).

Although most mathematics textbooks rely heavily on blocked practice, the Saxon series of mathematics textbooks include assignments with mostly interleaved practice (e.g., Saxon, 1997). Nevertheless, findings like the one reported here do not necessarily

suggest that Saxon texts produce higher test scores than do non-Saxon texts. This is because Saxon texts differ from non-Saxon texts in a number of ways other than the use of interleaved practice, and one or more of these other unique features may work against the efficacy of a Saxon text. In Saxon texts, for example, even the lessons are intermixed so that lessons on related topics (e.g., prism, cylinder, pyramid, and cone) do not appear consecutively—a feature that is not supported by the present study. Incidentally, Saxon texts have been criticized by some educators because of their purported emphasis on the mastery of procedures at the expense of conceptual understanding (e.g., Kamii & Dominick, 1998). However, we emphasize that this criticism of Saxon texts is orthogonal to the efficacy of interleaved practice. None of the authors of the present study have ever had any affiliation with Saxon textbooks.

Interleaved mathematics practice appears to be a feasible intervention. Creators of mathematics textbooks or instructional software need only rearrange practice problems before releasing the next edition, without altering the problems or lessons. Classroom implementation would also seem to require relatively little effort from teachers because they need not alter their classroom lessons or the manner in which they solve a particular problem. The intervention would also presumably require little or no teacher training, although teacher buy-in might require that teachers understand the logic underlying interleaved practice (namely, that it requires students to choose a strategy and not merely repeat the strategy used in the previous practice problem). The feasibility of an intervention also depends on whether students and teachers like it, yet little is known about the likability of interleaved practice. The few teachers in the present study indicated that they liked interleaved practice, even before the study was complete (i.e., before they learned of the intervention's efficacy), but their views might have been biased because of their close involvement with the project. Also, the present study did not include a survey of the students. Future research is needed to better gauge both students' and teachers' perceptions of interleaved practice.

Apart from its efficacy and feasibility, interleaved mathematics practice might be useful at all levels of mathematics, and this potential breadth contributes as much to the impact of an intervention as does its efficacy. Indeed, benefits of interleaved practice have been consistently observed with a variety of mathematics skills and with students in elementary school, middle school, and college. As argued here, these benefits arise because interleaved practice provides students with an opportunity to learn how to choose an appropriate strategy (or learn that they cannot do it). In short, interleaved practice simply provides students with an opportunity to practice the very skill they are expected to learn.

References

- Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, *120*, 20–33. doi:10.1037/0096-3445.120.1.20
- Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extent literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 83–105). London, England: Psychology Press.
- 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. doi:10.1037/0033-2909.132.3.354
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152. doi:10.1207/s15516709cog0502_2
- Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615–622. doi:10.1037/0021-9010.77.5.615
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58. doi:10.1177/1529100612453266
- Florida Department of Education. (2013). *2013 Florida statewide assessment scores*. Retrieved from <http://fcat.fldoe.org/mediapacket/2013/>
- Florida Department of Education, Bureau of K–12 Assessment. (2014). *Florida Comprehensive Assessment Test (Version 2.0)*. Tallahassee, FL: Author.
- Gay, L. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology*, *64*, 171–182. doi:10.1037/h0034595
- Grote, M. G. (1995). Distributed versus massed practice in high school physics. *School Science and Mathematics*, *95*, 97–101. doi:10.1111/j.1949-8594.1995.tb15736.x
- Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills*, *78*, 835–841. doi:10.2466/pms.1994.78.3.835
- Higgins, E., & Ross, B. (2011, November). *Comparisons in category learning*. Paper presented at the 52nd annual meeting of the Psychonomic Society, Seattle, WA.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*, 88–110. doi:10.1080/19345740802539325
- Kamii, C., & Dominick, A. (1998). The harmful effects of algorithms in Grades 1–4. In L. J. Morrow & M. J. Kenney (Eds.), *The teaching and learning of algorithms in school mathematics: 1998 NCTM yearbook* (pp. 130–140). Reston, VA: National Council of Teachers of Mathematics.
- Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2004). Timing of information presentation in learning statistics. *Instructional Science*, *32*, 233–252. doi:10.1023/B:TRUC.0000024191.27560.e3
- Küpper-Tetzel, C. E. (2014). Strong effects on weak theoretical grounds: Understanding the distributed practice effect. *Zeitschrift für Psychologie*, *222*, 71–81. doi:10.1027/2151-2604/a000168
- Le Blanc, K., & Simon, D. (2008). *Mixed practice enhances retention and JOL accuracy for mathematical skills*. Paper presented at the 49th annual meeting of the Psychonomic Society, Chicago, IL.
- Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, *35*, 105–123. doi:10.1901/jaba.2002.35-105
- National Center for Education Statistics. (2013). *The nation's report card: A first look. 2013 mathematics and reading* (NCES 2014–451). Washington, DC: U.S. Department of Education, Institute of Education Sciences.
- Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: Generation and interleaving effects. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the Cognitive Science Society* (pp. 1850–1855). Mahwah, NJ: Erlbaum.
- Richland, L. E., Linn, M. C., & Bjork, R. A. (2007). Cognition and instruction: Bridging laboratory and classroom settings. In F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 555–583). West Sussex, England: Wiley.
- Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory & Cognition*, *1*, 242–248. doi:10.1016/j.jarmac.2012.09.002
- Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*. Retrieved from <http://link.springer.com/article/10.3758%2Fs13423-014-0588-3>. doi:10.3758/s13423-014-0588-3
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209–1224. doi:10.1002/acp.1266
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems boosts learning. *Instructional Science*, *35*, 481–498. doi:10.1007/s11251-007-9015-8
- Saxon, J. H., Jr. (1997). *Algebra I* (3rd ed.). Norman, OK: Saxon.
- 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. doi:10.1111/j.1467-9280.1992.tb00029.x
- Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179–187. doi:10.1037/0278-7393.5.2.179
- Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. In J. Kilpatrick, G. W. Martin, & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 219–233). Reston, VA: National Council of Teachers of Mathematics.
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.
- Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science*, *30*, 759–774. doi:10.1207/s15516709cog0000_74
- Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*, 244–253. doi:10.3758/MC.38.2.244
- Taylor, K., & Rohrer, D. (2010). The effect of interleaving practice. *Applied Cognitive Psychology*, *24*, 837–848. doi:10.1002/acp.1598
- Yazdani, M. A., & Zebrowski, E. (2006). Spaced reinforcement: An effective approach to enhance the achievement in plane geometry. *Journal of Mathematical Sciences and Mathematics Education*, *1*, 37–43.

(Appendices follow)

Appendix A
Serial Position of Each Graph and Slope Problem in the Assignments

Students/Problem type	Practice assignment										Review	
	1	2	3	4	5	6	7	8	9	10		
Group 1												
Graph	1-4	—	1	7	4	6	10	1	10	10	6	
Slope	—	1-12	—	—	—	—	—	—	—	—	7	
Group 2												
Graph	1-12	—	—	—	—	—	—	—	—	—	6	
Slope	—	1-4	1	7	4	6	10	1	10	10	7	

Note. For example, for students in Group 1, Assignment 1 included four graph problems at the beginning of the assignment (Nos. 1–4), and Assignment 4 included one graph problem near the middle of the assignment (No. 7). The two groups received identical review assignments, and the graph and slope problems were the sixth and seventh problems, respectively. No student saw the same problem more than once during the study (including practice, review, and test).

Appendix B
Frequency of Responses of Three Teachers to Statements About Interleaved Practice

Statement	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Going over assignments was harder for me when the assignments were interleaved rather than blocked. (Neg)	2	1			
2. Going over assignments took more time when the assignments were interleaved rather than blocked. (Neg)	1	1		1	
3. Using interleaved assignments interfered with how I usually teach. (Neg)	2	1			
4. Students thought that interleaved assignments were more challenging than blocked assignments. (Neg)		1	1	1	
5. Students found that interleaved practice took longer than blocked practice. (Neg)		1	1	1	
6. Students liked interleaved practice less than blocked practice. (Neg)			2	1	
7. Interleaved assignments improved my students' learning more than did blocked practice.				3	
8. I would recommend interleaved practice to other math teachers.				3	
9. I would use interleaved practice in the future if it was an option.				2	1
10. I would use interleaved practice rather than blocked practice with lower-achieving students.		1	1	1	

Note. Statements regarding interleaved practice were framed negatively (Neg) in Statements 1–6.

Received May 13, 2014
 Revision received July 11, 2014
 Accepted July 11, 2014 ■