# A Randomized Controlled Trial of Interleaved Mathematics Practice

Doug Rohrer

Robert F. Dedrick

Marissa K. Hartwig

Chi-Ngai Cheung

University of South Florida

*Journal of Educational Psychology*

Published online May 24, 2019

Author Note

Doug Rohrer, Marissa K. Hartwig, and Chi-Ngai Cheung, Department of Psychology, University of South Florida; Robert F. Dedrick, Department of Educational and Psychological Studies, University of South Florida.

We thank Emily Gay and Harper Cassady for their help with scoring. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A160263 to the University of South Florida. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.

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

**Abstract**

We report the results of a pre-registered, cluster randomized controlled trial of a mathematics learning intervention known as interleaved practice. Whereas most mathematics assignments consist of a block of problems devoted to the same skill or concept, an interleaved assignment is arranged so that no two consecutive problems require the same strategy. Previous small-scale studies found that practice assignments with a greater proportion of interleaved practice produced higher test scores. In the present study, we assessed the efficacy and feasibility of interleaved practice in a naturalistic setting with a large, diverse sample. Each of 54 seventh-grade mathematics classes periodically completed interleaved or blocked assignments over a period of four months, and then both groups completed an interleaved review assignment. One month later, students took an unannounced test, and the interleaved group outscored the blocked group, 61% vs. 38%, $d = 0.83$. Teachers were able to implement the intervention without training, and they later expressed support for interleaved practice in an anonymous survey they completed before they knew the results of the study. Although important caveats remain, the results suggest that interleaved mathematics practice is effective and feasible.

## Educational Impact and Implications Statement

Every school day, many millions of mathematics students complete a set of practice problems that can be solved with the same strategy, such as adding fractions by finding a common denominator. In an alternative approach known as interleaved practice, practice problems are arranged so that no two consecutive problems can be solved by the same strategy, and this approach forces students to choose an appropriate strategy for each problem on the basis of the problem itself. We conducted a large randomized classroom study and found that a greater emphasis on interleaved practice dramatically improved test scores.

A typical mathematics assignment consists of a group of problems devoted to one skill or concept. For instance, a lesson on slope is usually followed by a set of a dozen or more slope problems, and this format is called *blocked practice*. Although a blocked assignment typically includes some kind of variety, such as a combination of procedural problems and word problems, every problem is nevertheless related to the same skill or concept. In an alternative approach known as *interleaved practice*, problems within an assignment are arranged so that no two consecutive problems require the same strategy, where strategy is defined loosely to include a procedure, formula, or concept. For example, a slope problem might follow a volume problem, and a probability problem about independent events might follow one about dependent events. Although blocked practice is more prevalent than interleaved practice, most students see both. For instance, students who ordinarily receive blocked assignments often see an interleaved review assignment before a cumulative exam. In the present study, each of 54 seventh-grade classes completed practice assignments that were either mostly blocked or mostly interleaved.

The study had two objectives. The first was to assess the efficacy of interleaved practice under naturalistic conditions. Most previous studies of interleaved practice have found that a greater emphasis on interleaving improved test scores, as we summarize further below, but these studies used small samples and some ecologically-invalid procedures (e.g., laboratory settings or only one session of practice). The present study examined interleaved practice in a large number of classes at multiple schools over a period of five months, and all instruction was delivered solely by teachers who had no prior association with the intervention or the authors. These kinds of realistic conditions are important because promising interventions often fizzle in the classroom (e.g., Hulleman & Cordray, 2009; O'Donnell, 2008).

The second objective was to evaluate the feasibility of implementing interleaved practice in classrooms – an issue not examined in previous studies. For instance, in order to assess whether teachers can incorporate interleaved practice in their courses, the teachers in the study

received no training or preparation. We also asked teachers to anonymously report their beliefs about interleaved practice because interventions sometimes fail without likability and teacher buy-in (e.g., Finn & Sladeczek, 2001).

**Blocked Practice**

Blocked practice appears to be far more common than interleaved practice, at least in the United States. In nearly every mathematics textbook we have examined, the majority of practice problems appear within blocked assignments. To be sure, most textbooks offer interleaved practice, usually in the form of review assignments described variously as Chapter Reviews, Mixed Reviews, or Spiral Reviews, but even these assignments often consist of several small blocks. For instance, most chapter reviews include several problems on the first lesson in the chapter, followed by several more on the second lesson, and so forth. The prevalence of blocked practice cannot be measured precisely, however, because an accurate census of adopted textbooks is not attainable. However, one formal evaluation of six seventh-grade mathematics textbooks found that, averaged across the texts, 78% of the practice problems were blocked, 11% were interleaved, and another 11% were difficult to classify (Dedrick, Rohrer, & Stershic, 2016). Moreover, blocked practice comprises 100% of the practice problems found in many consumable workbooks and internet-downloadable assignments, and these kinds of materials are increasingly supplementing or supplanting traditional textbooks.

Given the prevalence of blocked practice, one might reasonably wonder whether any evidence supports it. Specifically, once a student has worked several problems on the same skill or concept, is there any benefit of *immediately* working more problems of the same kind? Although this question has been asked by countless mathematics students, it has not received much attention from researchers. However, numerous studies of verbal learning have examined the effects of immediate, post-criterion practice. In these studies, subjects practiced a task until they reached a criterion of one correct response before either quitting or *immediately* continuing to practice the same task, and the subjects who continued to practice scored higher on a

subsequent test (e.g., Gilbert, 1957; Krueger, 1929; Postman, 1962; Rose, 1992). This effect was confirmed by meta-analysis (Driskell, Willis, & Cooper, 1992), although the same analysis revealed that the benefit of immediate, post-criterion practice rapidly diminishes after test delays exceeding one week. In brief, immediate post-criterion practice appears to improve the short-term learning of certain kinds of tasks, but we do not know of any such effects on mathematics learning.

Still, there are reasons to suspect that the blocking of similar practice problems might benefit learning. For instance, repeatedly solving problems of the same kind might reduce demands on working memory, and a number of studies have found that a concurrent working memory load (e.g., repeatedly rehearsing a seven-digit sequence) can impede performance on a variety of tasks, including puzzle solving (Kotovsky, Hayes, & Simon, 1985) and retrieval from long-term memory (Baddeley, Lewis, Eldridge, & Thomson, 1984). In fact, in numerous studies, mathematics practice problems were more effective when the problems were altered in ways that reduce cognitive load, which is akin to working memory load (e.g., Paas & Van Merriënboer, 1994; Sweller, 1994). In addition, blocked practice might benefit learning by reducing the number of students' errors, and, by the rationale underlying the strategy known as errorless learning, errors might impede learning by strengthening the association between a certain kind of problem and the incorrect solution (e.g., Skinner, 1958). However, this possibility is only speculative, and, moreover, several studies with non-mathematics tasks have found that students' errors can *enhance* their learning when errors are followed by corrective feedback (e.g., Huelser & Metcalfe, 2012; Kornell, Hays, & Bjork, 2009; Richland, Kornell, & Kao, 2009). In brief, there is some evidence to suggest that blocking practice problems might be advantageous, but the data are at best tangential.

Why, then, is blocked practice popular? One possibility is that students, teachers, and textbook authors might *believe* that blocking improves learning. With blocked practice, students know the strategy for each problem before they read the problem, and this resulting fluency,

though illusory, might lead students and teachers to falsely believe that blocking enhances efficacy (e.g., Koriat & Bjork, 2005; Kornell & Bjork, 2007; Schmidt & Bjork, 1992). Finally, and less provocatively, blocked practice might predominate textbooks simply because the authors find it convenient to follow each lesson with a group of problems devoted to that lesson.

**Interleaved Practice**

Although interleaved practice is much less common than blocked practice, there are good reasons to believe that interleaved practice enhances learning. Most notably, if an assignment includes a mixture of different kinds of problems, students cannot safely assume that a problem relates to the same skill or concept as does the previous problem, and thus the mixture provides students with an opportunity to choose an appropriate strategy on the basis of the problem itself, just as they must do when they encounter a problem on a cumulative exam. In effect, interleaved practice requires students to choose a strategy and not merely execute a strategy. This is not a trivial distinction because the choice of an appropriate strategy is often challenging (e.g., Siegler, 2003; Siegler & Shrager, 1984; Ziegler & Stern, 2014). This challenge is due partly to the sheer number of strategies from which students must choose, and partly to the fact that many problems lack features that clearly indicate which strategy is appropriate. For instance, a word problem that is solved by the Pythagorean Theorem might not include terms such as *hypotenuse* or *right triangle*, making it hard for students to infer that they should use the Pythagorean Theorem. In fact, students in nearly every mathematics discipline frequently encounter superficially-similar problems that require different strategies, forcing them to make fiendishly difficult discriminations (Table 1).

Insert Table 1 about here

In addition to any benefits of mixture per se, the interleaving of practice problems in a course or text inherently incorporates the learning strategies of spacing and retrieval practice, each of which is an effective and robust learning strategy.

**Spacing.** When practice problems within a textbook or course are rearranged to increase the degree of interleaving, the scheduling of each particular kind of problem is inherently distributed, or *spaced*, throughout the course to a greater degree. For instance, whereas most of the parabola problems in a mostly-blocked algebra textbook appear within a single assignment, most of the parabola problems in a mostly-interleaved textbook are distributed throughout the text. That is, when the practice of *multiple* skills is interleaved (ABCBACBCA) rather than blocked (AAABBBCCC), the practice of *any one* of the skills is necessarily spaced (A…A…A) rather than massed (AAA). In short, interleaved practice guarantees spaced practice.

Countless studies have found that a greater degree of spacing increased scores on a delayed test of learning, even when total time on task was equated (for a review, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). This spacing effect is large and robust, and it has been found with a wide variety of students, procedures, and learning materials. Spacing effects also have been found in a few studies of mathematics problem solving, including laboratory studies (Gay, 1973; Rohrer & Taylor, 2006, 2007), non-controlled classroom studies (Budé, Imbos, van de Wiel, & Berger, 2011; Yazdani & Zebrowski, 2006), and a randomized study embedded within a college mathematics course (Hopkins, Lyle, Hieb, & Ralston, 2016).

**Retrieval Practice.** With blocked practice, the formula or procedure needed to solve a problem is often the same as that needed to solve the previous problem, and this permits students to solve the problem without retrieving that information from memory. For example, if every problem in an assignment requires the same formula (slope = rise/run) or same procedure (find a common denominator), students need not retrieve this information from memory because they can instead obtain it by simply glancing at the solution to the previous problem, which is likely in plain view. With interleaved practice, however, the formula or procedure is not readily available, and thus students might first try to retrieve the information from memory before going to the trouble of finding the information or asking for help. In effect, the retrieval opportunity is an artifact of interleaved practice.

An attempt to retrieve information, when followed by feedback, is a learning strategy known as retrieval practice, and it has proven superior to other strategies (such as rereading the information) in many dozens of studies with verbal materials (Dunlosky et al., 2013; Roediger, Putnam, & Smith, 2011). Many of these studies had subjects learn paired associates such as HOUSE-CASA, and thus the benefits of retrieval practice probably extend to the learning of mathematical facts such as 8x5=40 or slope=rise/run (e.g., Pyke & LeFevre, 2011). Yet it is less clear whether retrieval practice enhances the *solving* of mathematics problems. A classroom-based study by Butler, Marsh, Slavinsky, and Baraniuk (2014) found a benefit of an intervention that combined retrieval practice and three other strategies, but a series of laboratory studies by Yeo and Fazio (in press) found mixed evidence for retrieval practice. In sum, whereas substantial evidence suggests that spacing improves mathematics problem solving, benefits of retrieval practice have yet to be demonstrated for mathematics tasks other than fact learning.

**Previous Studies of Interleaved Mathematics Practice**

There are multiple kinds of manipulations described as interleaving, and, in this literature review, we exclude studies of interventions that are fundamentally unlike the interleaving intervention assessed in the present study. Most notably, we have omitted studies of a well-known intervention in which every other practice problem is replaced by either a correct example (e.g., Sweller & Cooper, 1985; Van Gog & Lester, 2012) or an incorrect example (e.g., Booth, Lange, Koedinger, & Newton, 2013). Albeit supported by data, this alternation of example and practice problem nevertheless yields an assignment that is devoted to the same skill or concept (e.g., circumference), which means that the assignment is blocked. Thus, this kind of interleaving is the *complement* of the kind of interleaving manipulation examined in the present study. Our review also omits interleaving studies with category learning tasks, such as learning to classify statistical problems (Sana, Yan, & Kim, 2017) or chemical compounds (Eglington & Kang, 2017). In short, we focus here on studies in which students solved math

problems that were either interleaved or blocked by skill or concept. We know of 10 such studies.

The first four of these studies were conducted in laboratory settings. Mayfield and Chase (2002) had college students learn algebra rules with one of two methods that roughly correspond to interleaved and blocked practice (the study was designed for a different purpose), and they found a benefit of interleaved practice on a test given 4 – 12 weeks later. In a study by Rohrer and Taylor (2007), college students interleaved or blocked their practice of volume problems during two laboratory sessions one week apart, and interleaved practice improved test scores on a test given one week later (though interleaving worsened practice scores). This finding was replicated by Le Blanc and Simon (2008), who also explored issues unrelated to the efficacy of interleaved practice. Finally, in a study by Taylor and Rohrer (2010), fourth-grade students completed one session of interleaved or blocked practice of prism problems, and the interleaved group scored much higher on a test given one day later.

The remaining six studies took place in classroom settings. In a study with fifth- and sixth-grade students learning about fractions, Rau, Aleven, and Rummel (2013) found an interleaving benefit on tests given zero and seven days later (although the control group did not block practice in the usual sense). In two studies reported by Ziegler and Stern (2014), sixth-grade students saw two kinds of problems (addition and multiplication) that appeared either sequentially or juxtaposed (i.e., side by side), and juxtaposition led to better scores on tests given after delays of 1 day, 1 week, and 3 months. In similar studies reported by Rohrer, Dedrick, and Burgess (2014) and Rohrer, Dedrick, and Stershic (2015), seventh-grade students completed worksheets that provided a low or high dose of interleaved practice, and the heavier dose of interleaved practice led to higher scores on a test given after a delay of two weeks (in the first study) and delays of one day or one month (in the second study). Finally, in the largest previous study of interleaved practice (4 teachers and 146 students), Ostrow, Heffernan, Heffernan, and Peterson (2015) had seventh-grade students complete an interleaved or blocked

review assignment followed by a test 2 - 5 days later, and test scores showed a positive but not statistically significant effect of interleaving, though a post hoc median split analysis revealed a reliable interleaving benefit for the students with mathematics proficiency below the median. Altogether, these previous findings demonstrate that interleaved practice is a promising learning intervention that deserves greater scrutiny. Toward that aim, we designed the present study to evaluate the efficacy and feasibility of interleaved practice under naturalistic conditions with a large, diverse sample of students and teachers.

**The Present Study**

Each of 54 classes periodically received interleaved or blocked assignments over a period of four months before seeing an interleaved review assignment and an unannounced test one month later. Students received all instruction and assignments from their teachers, and the teachers had no prior association with the authors. We also took steps to prevent students and teachers from inferring the manipulation (see Method section). Unlike most previous studies, every student received an interleaved review assignment because many teachers provide such reviews before high-stakes tests, and thus the review ensured that the blocked practice condition was a realistic counterfactual (i.e., business as usual). Furthermore, the review assignment ensured that the time interval between the last practice problem of each kind (seen on the review assignment) and the test, an interval we call the *test delay*, was equated for both groups. Without the review, test delay would have been a confounding variable that worked in favor of the interleaved group.

Finally, although the experiment ostensibly compares interleaved and blocked practice, the manipulation is more accurately described as a comparison of mostly-interleaved and mostly-blocked practice because every student received *both* kinds of practice *outside* the experiment. For instance, we believe that all participating students received interleaved practice during their teachers' review for a district-required, semester exam (halfway through the practice phase) which covered every topic seen on the final test in the experiment. Likewise, students in both

groups almost certainly received some blocked practice (e.g., at least one worked example followed by at least a few practice problems) when their teachers first presented the skills and concepts covered in the experiment worksheets. In short, although we describe the groups as the interleaved group and blocked practice group, the experiment actually examined the efficacy of a low versus high dose of interleaved practice.

### Pilot Study

We conducted a pilot study at a public middle school in the school district where the main study took place. Two mathematics teachers participated, each with three classes of seventh grade students ($n$ = 83). Apart from sample size, the pilot study was nearly identical to the main study, and the minor procedural differences are noted in the Method section. On the test, the interleaved group ($M$ = 0.51, $SD$ = 0.32) outperformed the blocked group ($M$ = 0.22, $SD$ = 0.25). The effect size was large, $d$ = 0.97, 95% $CI$ = (0.52, 1.43).

## Method

The main study took place in a large school district in Florida during the 2017-2018 school year, one year after the pilot study. We preregistered the main study, and all materials and data are available at https://osf.io/pfeg4/. We received written permission from the university IRB, the school district, the principal of each participating school, each teacher, each student, and a parent of each student. The study was a cluster randomized controlled trial, with students nested within classes, and each class was randomly assigned to one of the two conditions.

### Participants

In order to determine the necessary number of participating classes, we conducted a priori power analyses with Optimal Design software (Raudenbush, Spybrook, Congdon, Liu, Martinez, Bloom, & Hill, 2011). Each analysis assumed a two-tailed test with an alpha level of .05 and a two-level, random effects model for a continuous outcome variable. We ran numerous analyses with varying values of effect size and intraclass correlation, all of which were more conservative than the values obtained in the pilot study. In every scenario, power exceeded .95 with 30

classes (15 per condition). We chose to recruit 50 classes, partly to allow for the attrition of teachers or schools, and partly because the marginal cost of each additional class was small in comparison to the cost of the entire study.

**Schools**. A school district official informed us that we could obtain our goal of 50 participating classes by recruiting five schools. We began with a list of the middle schools (grade 6-8) in the school district, and we excluded: 1) the school where we conducted the pilot study, 2) magnet and charter schools, 3) schools farther than a 30-min drive from the university, and 4) schools where fewer than 150 students passed the mathematics section of the sixth grade statewide assessment known as the Florida Standards Assessment (FSA) given in the spring of the previous school year. These criteria eliminated all but nine schools. We wrote the principals of these nine schools in a mostly serial fashion until we reached our goal of five participating schools. Ultimately, we wrote principals at only seven of the schools, two of whom did not respond to our e-mails. Each participating school received a $1000 donation.

**Teachers.** We recruited teachers who taught a seventh-grade math course described by the school district as Honors Advanced Grade 7 Mathematics. Although its title suggests that the course is selective, it is the modal course for seventh grade students at most of the schools in the district. The course excludes seventh-grade students enrolled in Algebra (one year earlier than most students in the district), and it excludes nearly all of the students who received a failing score (1 or 2 on a 5-point scale) on the mathematics section of the FSA taken at the end of the previous school year. More information about the student sample is given further below.

We recruited teachers who taught at least two sections of this course because our experimental design required that each teacher have at least one class in each condition. This within-teacher design enabled us to tease apart the teacher effect from the main effect of condition (e.g., Roberts, Lewis, Fall, & Vaughn, 2017). Although this design can lead to a kind of contamination known as treatment diffusion in which teachers use the intervention with students

in the control group, or vice-versa, we saw no evidence of this. At any rate, any treatment diffusion would have diminished the observed effect.

School administrators provided us with the names of 15 teachers who taught at least two sections of the selected course, and each of them agreed to participate in return for an honorarium of $1000. The 15 teachers (13 women and 2 men) were full-time middle school math teachers with a wide range of teaching experience (0 – 30+ years). None of the authors knew any of the teachers before the study began (but this was not true for the pilot study).

The participating classes were randomly assigned to either the interleaved or blocked condition with an algorithm that we ran for each teacher. For each teacher with 2, 4, or 6 classes, the algorithm evenly divided the classes into two groups by sampling without replacement, thereby ensuring that the teacher had the same number of classes in each condition (e.g., 2 and 2 rather than 3 and 1). For each teacher with 3 or 5 classes, the algorithm first randomly assigned one class to a randomly-chosen condition before evenly dividing the remaining classes by sampling without replacement. Ultimately, this algorithm assigned 28 classes to the interleaved condition and 26 classes to the blocked condition. The breakdown for each teacher is shown in Table 2.

Insert Table 2 about here

**Students.** We began recruiting students in September. Students were told that we were seeking permission to use their solutions to math problems for a research study in return for a $20 gift card. When we began recruiting, the participating classes included 1103 students. Of these students, 21 students (2%) returned consent forms with a decline response from the student or parent, 226 students (20%) returned no forms or incomplete forms, and 856 students (78%) agreed to participate by providing both their written assent and their parent's written permission. Of these 856 students who began the study, 69 students (8%) either withdrew from their course during the study or did not attend class on the day of the unannounced test. The attrition rate was about the same for the interleaved group (39/437 = 8.9%) and the blocked

group (30/419 = 7.2%). Thus, the final sample included 787 students, and only their test scores were analyzed. Table 2 shows the nesting of students within classes within teachers within schools.

After we completed the study, the school district provided us with additional data for the participating students, aggregated by condition. These data included students' score on the mathematics section of their Grade 6 FSA, and this measure showed no significant difference between the interleaved group ($M$ = 345, $SD$ = 12) and the blocked group ($M$ = 346, $SD$ = 12). For this test, the range of possible scores is 260-390, and the state-mandated passing score is 325. The school district also provided demographic measures such as sex and race, and we found no reliable differences between the two groups on these measures either (Table 3).

Insert Table 3 about here

**Timeline**

The study included three parts: a practice phase with eight worksheets, a review worksheet, and a test. The entire procedure lasted about five months, and the time course varied slightly across teachers. Averaged across classes, the practice phase (Worksheets 1-8) spanned 103 days (range 98-108 days), followed by the review assignment 10 days later (range 7-14 days), which in turn was followed by the test 33 days later (range 28-40 days). Although these time intervals varied across teacher, time interval is not a confounding variable because each teacher had classes in both conditions. In the pilot study, the practice phase spanned 47 days, followed 4 or 5 days later by the review, followed 30 days later by the test. We did not administer a pretest, primarily because we were reluctant to ask teachers to sacrifice a class meeting early in the school year, before we had established a rapport.

**Worksheets**

We created every problem. We first wrote a much larger set of problems and then revised or omitted problems on the basis of feedback we received during several meetings with two highly-

experienced middle school mathematics teachers who participated in the pilot study. Each problem required a concrete solution (e.g., no open-ended items).

The worksheets included critical problems and filler problems. The critical problems were like the kinds of problems seen on the test, and these consisted of four kinds: Graph (A), Inequality (B), Expression (C), and Circle (D). An example of each is shown in Figure 1. The filler problems were drawn from topics unrelated to the critical problems (e.g., probability, angles, volume), and we included filler problems partly to prevent students and teachers from inferring the difference between the two conditions. Specifically, for the interleaved group, the critical problems were interleaved, yet many filler problems were blocked. Similarly, for the blocked group, the critical problems were blocked, yet most filler problems were interleaved. For students and teachers, the interleaved and blocked conditions were known simply as the green and blue conditions, respectively.

Insert Figure 1 about here

The arrangement of problems on each worksheet is shown in Figure 2. Several features warrant mention. 1) Each worksheet had eight practice problems. 2) Every student saw the same practice problems though not in the same order, and no student saw the same problem twice. 3) For each kind of critical problem (A, B, C, or D), the practice problems appeared in the same order. Thus, the first circle problem seen in the blocked condition was identical to the first circle problem seen in the interleaved condition. 4) Worksheet 9 (the review) was the same for both groups, and it included one of each kind of critical problem, thereby ensuring that the test delay for each kind of critical problem was the same in both conditions. 5) For the interleaved group, the critical problems on Worksheets 1-8 were arranged so that each kind of critical problem immediately preceded each one of the other kinds equally often. 6) Although Worksheet 9 served as a review, neither students nor teachers were told that it was a review. Thus, the students presumably believed that Worksheet 9 was merely another worksheet, although the teachers likely noticed that this worksheet was the same for both conditions.

Insert Figure 2 about here

Each worksheet spanned two sides of a single sheet of paper. Interleaved worksheets were printed on green paper, and blocked worksheets were printed on blue paper. Each answer key was printed on white paper, and teachers received a separate answer key for each class. The worksheets and answer key for each participating class were placed in a large plastic envelope labeled with the teacher name and class period. We hand-delivered the envelopes to teachers at their schools and, during a subsequent visit, collected the envelopes with the completed copies. We always collected Worksheet N before giving teachers Worksheet N+1.

Students completed the worksheets during class under the supervision of their teachers. The teachers received the following paraphrased instructions: 1) Begin the activity with at least 30 minutes remaining in the class period. 2) Have students work on the problems until nearly all students are finished or no longer making progress. 3) If you wish, you may provide one-on-one help to students while they work on the problems. 4) Once most students finish or stop making progress, place the answer key on your document camera and present each solution one at a time. 5) For each solution, give students an opportunity to ask questions, and ask students to correct any errors in their answers and solutions.

**Fidelity**

Treatment fidelity was good. Every teacher distributed each of the nine worksheets to each of their participating classes, and our one-at-a-time delivery procedure ensured that the teachers presented these worksheets in the specified order. However, we know of instances in which teachers did not follow instructions. On several occasions, teachers did not allot enough class time for a worksheet, and their students did not finish the worksheet until the next class meeting. Also, at least one teacher did not present the answer key with the document camera and instead had some students write the solutions on a white board. When we learned of such behaviors during our periodic school visits, we reminded teachers of the protocol.

Student compliance was generally strong. We received all nine worksheets for 61% of the students, at least seven worksheets for 98% of the students, and at least five worksheets for every student. Details are given in Table 4. The four authors and two research assistants scored every problem on every worksheet we received from students, which totaled more than 50,000 problems. For these problems, students provided the correct (or corrected) answer for 95% of the problems, and further details are provided in Table 5. Notably, these worksheet scores provide a measure of compliance, not performance, partly because students were allowed to seek help while they tried to solve the practice problems, and partly because students were asked to correct their errors once they saw the solutions. Thus, we have no measure of students' performance on the practice problems.

Insert Tables 4 and 5 about here

**Test**

We tested students on five days in March – a different date for each school. Students were tested during their regular class meeting in the presence of their teacher and at least one author. Students who were absent on their assigned test day did not take the test. We asked teachers not to inform students of the test in advance, and teachers received no information about the test content in advance.

Each test booklet included a cover sheet and four test pages, each printed on one side only. The test included four graph problems (page 1), four inequality problems (page 2), four expression problems (page 3), and four circle problems (page 4). We chose to block the test problems because some researchers have suggested that an interleaving test format would favor students who interleaved their practice, and, if this is true, our choice of a blocked format would have worked *against* an interleaving benefit. We chose the sequence of these blocks (graph problems, then inequality problems, then expression problems, and then circle problems) because we believe it ordered the four kinds of problems from least to most time-consuming. None of the test problems had appeared previously in the study. Every student saw the same

test problems, but we created four versions of the test by reordering the problems within each page. An author distributed every test booklet to students and ensured that adjacent students received different test versions. In addition, teachers separated students' desks or required students to use dividers that ostensibly prevented them from seeing other students' tests. Students were allotted 25 minutes and allowed to use a calculator.

Every test was scored at the school on the day of the test by the four authors and two research assistants. Scorers were blind to condition, and each answer was marked as correct or not. Two scorers independently scored each test. Discrepancies were rare (83 in 12,592), and the four authors later met and resolved each discrepancy. The internal consistency reliability of the test was high (for the 16 items, Cronbach's alpha = .89).

**Teacher Survey**

Several weeks after the test, the second author hand-delivered to each teacher a 23-item paper-and-pencil survey and a stamped envelope addressed to the author. Each teacher was asked to anonymously complete the survey and return it by mail. All teachers returned the survey, and only then did we inform them of the results and purpose of the study. The survey items were preceded by a brief tutorial about interleaved and blocked practice. The survey appears in the Appendix.

## Results

On the test, the interleaved group outscored the blocked group by a large margin. Table 6 lists descriptive measures. The effect size was large, Cohen's *d* = 0.83, 95% CI = [0.68, 0.97], where $d = (M_1 - M_2)/SD_{pooled}$, and *M* and *SD* are based on the student-level data (not class means). We observed a positive effect for each of the 15 teachers, *d*s = 0.23 – 1.48.

Insert Table 6 about here

We also found a positive interleaving effect for each of the four kinds of critical problems (A, B, C, and D), but the effect sizes are misleading. Ranked from largest to smallest, the Cohen's *d* values for the four kinds equaled 0.86 (A), 0.63 (B), 0.40 (C), and 0.34 (D), and this rank order

corresponds to the order in which the blocked group saw these kinds of problems during the practice phase – not coincidentally, we believe. That is, the largest effect was observed for *A* problems (graphs), but the blocked group worked the *A* problems early in the practice phase, long before the test, thereby disadvantaging the blocked group and thus inflating the interleaving effect. By contrast, the *D* problems (circles) produced the smallest interleaving effect, yet the blocked group worked the *D* problems near the *end* of the practice phase, which shortened their test delay, and the shorter test delay likely boosted their test scores and thus dampened the interleaving effect. In brief, these unavoidable scheduling confounds likely contributed to the large differences in the effect sizes of the four kinds of critical problems. However, there was no such confound for the critical problems as a whole because, when average across all four kinds, the time interval between each practice problem and the test date was nearly equal for the two groups (the slight difference favored the blocked condition).

**Multilevel Modeling Analysis**

Because of the cluster design, we further examined test scores by fitting a two-level model (students within classes) with HLM Version 7.03 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). Using restricted maximum likelihood (REML), we first estimated a fully unconditional model to evaluate the variability in students' scores within and between classes (Table 7). To assess the difference between conditions, we used REML to estimate a two-level random-intercept model. Tests of the distributional assumptions about the errors at each level of the model (normality and equal variance) did not reveal any violations. The level-2 class model included a dummy variable for condition (0=Blocked, 1=Interleaved) and 14 dummy variables for teacher effects. Before examining the main effect of condition, we evaluated the potential interaction between teacher and condition and found no statistically significant interaction effects, $p > .05$. We then tested a main effects model that evaluated the effect of condition, controlling for teacher effects, and we found a significant effect of interleaving ($p < .001$). Details are provided in Table 7.

Insert Table 7 about here

**Teacher Survey**

The results of the anonymous teacher survey are shown in the Appendix, and here we briefly summarize the results. A majority of the 15 teachers indicated that their students found interleaved practice to be slightly harder (9) and slightly more time-consuming (13) than blocked practice. Most also agreed that presenting the solutions to an interleaved assignment took slightly more time than it did for a blocked assignment (8), yet they reported that the difficulty of doing so was about the same (11). Some reported that their students disliked blocked practice more than interleaved practice (6), and others indicated that student likability was about the same (5).

For all other items, interleaved practice was judged favorably. Teachers agreed (or strongly agreed) that interleaved practice is a good way to improve students' scores on unit exams (14) and final exams (14) and is appropriate for both low-achieving students (13) and high-achieving math students (15). Most teachers also agreed (or strongly agreed) that they could give interleaved assignments without changing how they ordinarily teach (11), and they wished that their students' instructional materials included more interleaved practice (12). Most also reported that they liked interleaved practice (13) and that they would recommend it to other math teachers (13). Finally, most agreed that other math teachers would be willing to use interleaved practice (11) and would be able to do so with little or no instruction (12). In summary, most of the teachers reported that interleaved practice was useful and viable, yet a majority reported that their students found interleaved practice to be “slightly” harder and more time-consuming than blocked practice.

## Discussion

In the large-scale randomized control trial presented here, a higher dose of interleaved practice increased scores on a delayed, unannounced test. The effect size was large, and a positive effect was found for each of the 15 teachers. This finding is consistent with the results

of previous small-scale studies of interleaved mathematics that found test benefits with a variety of materials, procedures, and students. Taken as a whole, the extant evidence suggests that interleaved mathematics practice is effective and robust, though we list several caveats below.

The effect size observed in the present study might seem surprisingly large for a classroom-based experiment, but this might be due to fact that interleaved mathematics practice combines three potent learning strategies, as explained in the Introduction. First, the mixture of different kinds of problems *within* each assignment provides students with an opportunity to practice choosing a strategy on the basis of the problem itself, which is precisely what students must do when they encounter a problem on a cumulative exam or other high-stakes test. Second, interleaved mathematics practice inherently ensures a greater degree of spaced practice of each particular skill or concept *across* assignments, allowing students to exploit the spacing effect. Third, interleaving might encourage students to engage in the strategy known as retrieval practice by leading them to recall, or at least try to recall, the information needed to solve the problem (e.g., slope = rise / run). The secondary benefits of spacing and retrieval practice are not trivial. In one commissioned evaluation of 10 learning strategies, spacing and retrieval practice were the only strategies to receive the highest possible rating (Dunlosky et al., 2013).

**Caveats**

Although the present study found a large effect of interleaved practice, the effect size likely depends on other factors. This list includes the usual possibilities, such as student proficiency, teacher buy-in, duration of the intervention, choice of material, and degree of transfer required by the outcome measure. Apart from these possible moderators, there are four caveats that we believe might be crucial.

1. Interleaved practice probably takes more time, which is to say that students need more time to complete a particular practice problem when it is part of an interleaved assignment rather than a blocked assignment. Although we did not measure students' time on task, every teacher reported that the interleaved assignments took more time than did blocked practice. To

the extent that this was true, the observed interleaving effect would have been smaller if it had been measured per unit of *time* invested by the student. To our knowledge, no previous study of interleaved mathematics practice has measured time on task, which probably requires computer-based data collection.

2. The test benefit of interleaved mathematics practice might be smaller at shorter test delays. In fact, in the one previous interleaved mathematics study that included a manipulation of test delay, the interleaving effect was smaller at the shorter test delay (Rohrer et al., 2015). Furthermore, the only previous study that did not find a positive interleaving effect used a relatively brief test delay of 2 - 5 days (Ostrow et al., 2015), although some of the other studies found positive interleaving effects after test delays of one day or less (e.g., Taylor & Rohrer, 2010).

3. Interleaved practice might be less effective or too difficult if students do not first receive at least a small amount of blocked practice when they encounter a new skill or concept. As explained in the Introduction, the interleaved group in the present study likely received at least some blocked practice from their teachers *before* they received the experiment worksheets, and this appears to be true for the other math interleaving studies with one exception (Rohrer & Taylor, 2007). In brief, the data do not suggest that students entirely avoid blocked practice.

4. Interleaved practice might be effective only if students receive corrective feedback. The students in the present study were shown the solutions and asked to correct their errors, and it appears that feedback also was provided to students in every previous study of interleaved mathematics practice (see Introduction). Thus, informative and timely feedback might be a necessary ingredient of interleaved practice.

**Feasibility**

The results of the present study also suggest that interleaved mathematics practice can be feasibly implemented in the classroom. The participating teachers were able to incorporate interleaved practice in their classrooms without training or support, and most reported that the

intervention is effective and easy to use. Nearly all of them also reported that interleaved practice is appropriate for both low- and high-achieving students.

However, we do not know students' beliefs about interleaved practice. Although teachers in the present study reported that their students found interleaved and blocked practice to be about equally likeable, we did not ask the students for their views. Future research might also examine whether students *believe* that interleaved practice is effective because students who doubt its utility might be less likely to use it. These kinds of metacognitive beliefs have been surveyed for some learning strategies (e.g., Hartwig & Dunlosky, 2012) but not for interleaved mathematics practice. However, in scenarios involving non-mathematics category learning tasks, previous studies have found that a majority of students mistakenly believed that blocked practice is more effective than interleaved practice (Kornell & Bjork, 2008; McCabe, 2011).

In our view, the greatest barrier to the classroom implementation of interleaved mathematics practice is the relative scarcity of interleaved assignments in most textbooks and workbooks. There are some remedies, though. For instance, teachers can create interleaved assignments by simply choosing one problem from each of a dozen assignments from their students' textbook (such as Problem #6 on p. 45, Problem #12 on p. 33, and so forth). Teachers might also search the Internet for worksheets providing "mixed review" or "spiral review," and they can use practice tests created by organizations that create high-stakes mathematics tests. Ultimately, though, we hope that the publishers of textbooks, workbooks, and instructional software add more interleaved practice to their products. These materials are typically updated every few years, and, as part of this revision, a portion of the blocked practice in the previous edition can be replaced by interleaved practice. This route of implementation is not particularly novel. Creators of learning materials have often incorporated recommendations by researchers and educational organizations when updating their materials, and doing so is in their financial interest.

**Final Thought**

The present study provides another illustration of how a simple and inexpensive intervention can improve learning. While many unproven and expensive educational products continue to garner media attention and tax dollars, numerous classroom-based randomized experiments have found benefits of straightforward interventions requiring neither technology nor proprietary materials (Roediger & Pyc, 2012). For instance, Ramani, Siegler, and Hitti (2012) found that playing a simple board game improved preschoolers' understanding of number magnitude, and McNeil, Fyfe, and Dunwiddie (2015) found that minor reformatting of arithmetic problems improved second graders' understanding of mathematical equivalence (2 + 7 = 6 + _ ). These kinds of studies demonstrate that an intervention can be effective without being flashy, and we hope that the present study contributes to a greater appreciation of the difference.

## References

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, *113*(4), 518–540.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25,* 24–34*.*

Budé, L., Imbos, T., van de Wiel, M. W., & Berger, M. P. (2011). The effect of distributed practice on students' conceptual understanding of statistics. *Higher Education, 62*, 69–79.

Butler, A. C., Marsh, E. J., Slavinsky, J. P., & Baraniuk, R. G. (2014). Integrating cognitive science and technology improves learning in a STEM classroom. *Educational Psychology Review*, *26*(2), 331–340.

Dedrick, R. F., Rohrer, D., & Stershic, S. (2016, April). *Content analysis of practice problems in 7th grade mathematics textbooks: Blocked vs. interleaved practice.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

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

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

Eglington, L. G., & Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition, 6,* 475–485*.*

Finn, C. A., & Sladeczek, I. E. (2001). Assessing the social validity of behavioral interventions: A review of treatment acceptability measures. *School Psychology Quarterly, 16*(2), 176-206.

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

Gilbert, T. F. (1957). Overlearning and the retention of meaningful prose. *Journal of General Psychology, 56*, 281–289.

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

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. (2016). Spaced retrieval practice increases college students' short-and long-term retention of mathematics knowledge. *Educational Psychology Review*, *28*(4), 853–873.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514–527.

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

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187–194.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219–224.

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

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248–294.

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

Le Blanc, K., & Simon, D. (2008, November). *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.

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

McNeil, N. M., Fyfe, E. R., & Dunwiddie, A. E. (2015). Arithmetic practice can be modified to promote understanding of mathematical equivalence. *Journal of Educational Psychology*, *107*(2), 423–436.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*, 33–84.

Ostrow, K., Heffernan, N., Heffernan, C., & Peterson, Z. (2015). Blocking vs. interleaving: Examining single-session effects within middle school math homework. In *Artificial Intelligence in Education* (pp. 338–347). Springer International Publishing.

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

Postman, L. (1962). Retention as a function of degree of overlearning. *Science, 135*, 666–667.

Pyke, A. A., & LeFevre, J. A. (2011). Calculator use need not undermine direct-access ability: The roles of retrieval, calculation, and calculator use in the acquisition of arithmetic facts. *Journal of Educational Psychology*, *103*(3), 607-616.

Ramani, G. B., Siegler, R. S., and Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology*, *104*(3): 661–672.

Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction*, *23*, 98–114.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7*: *Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., Martinez, A., Bloom, H., & Hill, C. (2011). Optimal Design plus empirical evidence (Version 3.0).

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243–257.

Roberts, G., Lewis, N. S., Fall, A. M., & Vaughn, S. (2017). Implementation fidelity: Examples from the reading for understanding initiative. In G. Roberts, S. Vaughn, S. N. Beretvas, & V. Wong (Eds.), *Treatment fidelity in studies of educational intervention* (pp. 61–79). New York, NY: Routledge.

Roediger, H. L., Putnam, A. L. & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 1–36). Oxford: Elsevier.

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 and Cognition*, *1*(4), 242–248. http://dx.doi.org/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, 21,* 1323–1330.

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*, 900–908

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.

Rose, R. J. (1992). Degree of learning, interpolated tests, and rate of forgetting. *Memory & Cognition, 20*, 621–632.

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*(1), 84–98. doi:10.1037/edu0000119

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.

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.

Skinner, B. F. (1958). Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science, 128,* 969–977.

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

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

Van Gog, T., & Kester, L. (2012). A test of the testing effect: acquiring problem-solving skills from worked examples. *Cognitive Science*, *36*(8), 1532–1541.

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.

Yeo, D. J., & Fazio, L. K. (in press). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *Journal of Educational Psychology.* http://dx.doi.org/10.1037/edu0000268

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction, 33,* 131–146.

## Appendix

Teacher Questionnaire

Now that you and your students have completed the research study, we would like to know your opinions about the assignments. As you probably noticed, the research study included two kinds of assignments.

In each blocked assignment, every problem was related to the same concept or procedure. For example, one of the blocked assignments included only pyramid problems. In each interleaved assignment, no two problems were related to the same concept or procedure. For example, one of the interleaved assignments included one circle problem, one triangle problem, and so forth.

Please answer the questions on these pages. There are no wrong answers. Feel free to skip a question. When you are finished, place this page in the enclosed envelope. Do not write your name on this paper.

| | Interleaved Assignments ... | Interleaved Assignments ... | About the Same | Blocked Assignments ... | Blocked Assignments ... |
|---|---|---|---|---|---|
| 1. Which kind of assignment took students more time to finish? | Took Much More Time<br>2 | Took Slightly More Time<br>13 | 0 | Took Slightly More Time<br>0 | Took Much More Time<br>0 |
| 2. Which kind of assignment was harder for students? | Were Much Harder<br>1 | Were Slightly Harder<br>9 | 2 | Were Slightly Harder<br>3 | Were Much Harder<br>0 |
| 3. Which kind of assignment took you more time to go over? | Took Much More Time<br>2 | Took Slightly More Time<br>8 | 4 | Took Slightly More Time<br>1 | Took Much More Time<br>0 |
| 4. Which kind of assignment was harder for you to go over? | Were Much Harder<br>0 | Were Slightly Harder<br>3 | 11 | Were Slightly Harder<br>0 | Were Much Harder<br>0 |
| 5. Which kind of assignment did students dislike more? | Were Disliked Much More<br>0 | Were Disliked Slightly More<br>3 | 5 | Were Disliked Slightly More<br>5 | Were Disliked Much More<br>1 |

| | Strongly Disagree | Disagree | Neither Agree Nor Disagree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 6. I understand what interleaved practice is. | 0 | 0 | 0 | 3 | 12 |
| 7. I understand the logic of interleaved practice. | 0 | 0 | 0 | 6 | 9 |
| 8. Interleaved practice is a good way to improve students' scores on unit exams. | 0 | 0 | 1 | 6 | 8 |
| 9. Interleaved practice is a good way to improve students' scores on final exams. | 0 | 0 | 1 | 6 | 8 |
| 10. Interleaved practice is appropriate for high-achieving math students. | 0 | 0 | 0 | 3 | 12 |
| 11. Interleaved practice is appropriate for low-achieving math students. | 0 | 0 | 2 | 9 | 4 |
| 12. I could give interleaved assignments without changing how I ordinarily teach. | 0 | 2 | 2 | 6 | 5 |
| 13. I wish my students' workbook or textbook included more interleaved assignments. | 1 | 0 | 2 | 5 | 7 |
| 14. I could easily create my own interleaved assignments. | 0 | 5 | 1 | 5 | 4 |
| 15. I would recommend interleaved practice to other math teachers. | 0 | 0 | 2 | 7 | 6 |
| 16. Most math teachers would be willing to use interleaved practice in their classroom. | 0 | 0 | 4 | 7 | 4 |
| 17. Most math teachers could learn to use interleaved practice in their class with little or no instruction. | 0 | 0 | 3 | 8 | 4 |
| 18. I like interleaved practice. | 0 | 0 | 2 | 4 | 9 |

*Note*. One teacher did not respond to Item 4, and another teacher did not respond to Item 5. For brevity, the wording of questions 1-5 shown here differed slightly from the original version. The original version is posted on OSF.

**Table 1**

*Superficially-Similar Mathematics Problems That Require Different Strategies*

| Problem | Strategy |
|---|---|
| Algebra | |
| Solve. $x - 4x + 3 = 0$ | Group x terms on one side |
| Solve. $x^2 - 4x + 3 = 0$ | Factor or quadratic formula |
| Geometry | |
| Find the length of the line segment with endpoints (1, 2) and (5, 5) | Pythagorean Theorem |
| Find the slope of the line segment with endpoints (1, 2) and (5, 5) | Rise / Run |
| Trigonometry | |
| For $\triangle XYZ$, find *x* if $\angle X = 60°$, y = 3, and z = 5. | Law of Cosines |
| For $\triangle XYZ$, find *x* if $\angle X = 60°$, y = 3, and $\angle Y = 50°$. | Law of Sines |
| Calculus | |
| $\int x(e + 1)^x \, dx$ | Integration by Parts |
| $\int e(x + 1)^e \, dx$ | U-Substitution |

**Table 2**

*Participant Nesting*

| | | Interleaved | | Blocked | |
|---|---|---|---|---|---|
| School | Teacher | Number of Classes | Number of Students | Number of Classes | Number of students |
| A | 1 | 2 | 11 | 2 | 39 |
| | 2 | 2 | 36 | 2 | 25 |
| | 3 | 2 | 25 | 2 | 21 |
| B | 4 | 1 | 15 | 1 | 17 |
| | 5 | 1 | 15 | 1 | 29 |
| | 6 | 1 | 14 | 1 | 17 |
| C | 7 | 2 | 24 | 2 | 27 |
| | 8 | 2 | 10 | 2 | 15 |
| D | 9 | 3 | 42 | 2 | 24 |
| | 10 | 2 | 35 | 1 | 17 |
| | 11 | 2 | 38 | 2 | 38 |
| | 12 | 1 | 16 | 1 | 12 |
| E | 13 | 2 | 29 | 2 | 30 |
| | 14 | 3 | 54 | 3 | 51 |
| | 15 | 2 | 34 | 2 | 27 |
| Total | | 28 | 398 | 26 | 389 |

**Table 3**

*Student Demographics*

| | Interleaved (*n* = 398) | | Blocked (*n* = 389) | | All (*n* = 787) | |
|---|---|---|---|---|---|---|
| | *n* | *%* | *n* | *%* | *n* | *%* |
| Sex | | | | | | |
| Female | 212 | 53.3 | 207 | 53.2 | 419 | 53.2 |
| Male | 186 | 46.7 | 182 | 46.8 | 368 | 46.8 |
| Race | | | | | | |
| Asian | 17 | 4.3 | 29 | 7.5 | 46 | 5.8 |
| Black | 31 | 7.8 | 28 | 7.2 | 59 | 7.5 |
| Hispanic | 75 | 18.8 | 75 | 19.3 | 150 | 19.1 |
| White | 254 | 63.8 | 234 | 60.2 | 488 | 62.0 |
| Other | 21 | 5.3 | 23 | 5.9 | 44 | 5.6 |
| FRL | 105 | 26.4 | 84 | 21.6 | 189 | 24.0 |
| ELL/LEP | 8 | 2.0 | 10 | 2.6 | 18 | 2.3 |

*Note*. FRL = Free/Reduced Lunch; ELL = English Language Learner; LEP = Low English Proficiency. The school district did not provide student ages, but most seventh-grade students in the district are 12 years of age at the beginning of the school year.

**Table 4**

*Worksheets Received, as a Percentage of the Number of Students*

| | | Worksheet | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | *n* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean |
| Interleaved | 398 | 95.2 | 93.5 | 95.0 | 89.7 | 95.0 | 96.5 | 92.2 | 92.2 | 91.7 | 93.4 |
| Blocked | 389 | 94.6 | 94.3 | 96.1 | 91.8 | 95.4 | 95.6 | 93.3 | 93.6 | 94.6 | 94.4 |
| All | 787 | 94.9 | 93.9 | 95.6 | 90.7 | 95.2 | 96.1 | 92.8 | 93.6 | 93.1 | 94.0 |

*Note*. Worksheet 9 was the review worksheet.

**Table 5**

*Percentage of Practice Problems with Correct Answers*

| | | Worksheet | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean |
| Interleaved | *M* | 94.1 | 94.2 | 95.5 | 96.1 | 94.9 | 94.3 | 96.3 | 97.3 | 97.0 | 95.5 |
| | *SD* | 11.3 | 11.7 | 10.6 | 8.0 | 9.7 | 10.6 | 9.9 | 6.9 | 8.1 | 9.6 |
| Blocked | *M* | 96.2 | 88.1 | 96.4 | 92.3 | 97.0 | 95.3 | 97.1 | 96.3 | 95.4 | 94.9 |
| | *SD* | 10.0 | 23.2 | 9.9 | 18.6 | 8.0 | 16.6 | 8.0 | 14.3 | 10.7 | 13.3 |
| All | *M* | 95.1 | 91.2 | 96.0 | 94.2 | 95.9 | 94.8 | 96.7 | 96.8 | 96.2 | 95.2 |
| | *SD* | 10.7 | 18.6 | 10.3 | 14.4 | 8.9 | 13.9 | 9.0 | 11.2 | 9.5 | 11.8 |

*Note*. Each percentage is based on the total number of problems appearing on the worksheets we received from teachers (see Table 4). Worksheet 9 was the review worksheet. These scores represent a measure of compliance, not performance, because worksheets were scored *after* students were shown the solutions and asked to correct their errors.

**Table 6**

*Descriptive Statistics for the Test (% correct)*

| | Interleaved (*n* = 398) | Blocked (*n* = 389) | Total (*n* = 787) |
|---|---|---|---|
| *M* | 60.7 | 37.6 | 49.3 |
| *SD* | 28.6 | 27.3 | 30.3 |
| Median | 62.5 | 31.3 | 50.0 |
| Range | 0-100 | 0-100 | 0-100 |
| Skewness | -0.33 | 0.45 | 0.07 |
| Kurtosis | -1.03 | -0.68 | -1.15 |

**Table 7**

*Two-Level Model of Test Score (% correct)*

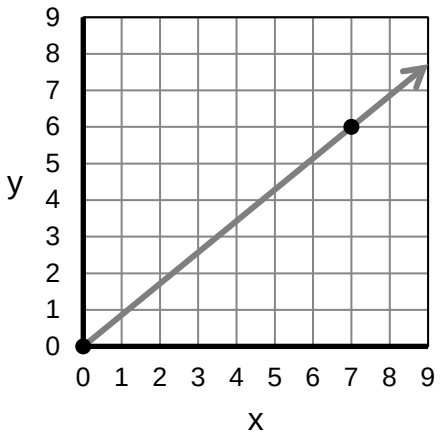
| | Model 1 | Model 2 |
|---|---|---|
| Fixed effects | | |
| Intercept | 49.48 (2.20) *** | 24.55 (4.57) *** |
| Interleaved Practice | | 22.17 (2.47) *** |
| | | |
| Variance components | | |
| Between classroom | 206.72*** | 30.13* |
| Within classroom | 694.13 | 693.68 |

*Note*. Parenthetical values are standard errors. Model 1 is an unconditional model. Model 2 included a dummy variable for condition (0 = Blocked, 1 = Interleaved) and 14 dummy variables for the 15 teachers. Intraclass correlation from Model 1 equals = .23. Tests of significance of the within-classroom variance are not conducted in HLM Version 7.03.

*$p < .05$. **$p < .01$. ***$p < .001$.

**Figure 1**

**A** Write an equation of the form $y = kx$ for the proportional relationship shown in the graph.




$$k = \frac{y}{x} = \frac{6}{7}$$

$$y = \frac{6}{7}x$$

**B** Solve the inequality.

$$-5x + 5 > -40$$
$$-5 \quad -5$$

$$\frac{-5x}{-5} > \frac{-45}{-5}$$

$$x < 9$$

**C** Simplify the expression.

$$4(-2x - 1) - 3(-5x - 2)$$

$$= -8x - 4 + 15x + 6$$

$$= 7x + 2$$

**D** A circle has an area of 254.34 square cm. Find its radius, in cm. Use $\pi = 3.14$

$$A = \pi r^2$$

$$\frac{254.34}{3.14} = \frac{3.14\, r^2}{3.14}$$

$$81 = r^2$$

$$9 = r$$

*Figure 1*. The four kinds of problems appearing on the test. Students saw graph problems (A), inequalities (B), expressions (C), and circles (D). The solutions shown above are identical to the ones appearing on the answer key shown to students.

**Figure 2**

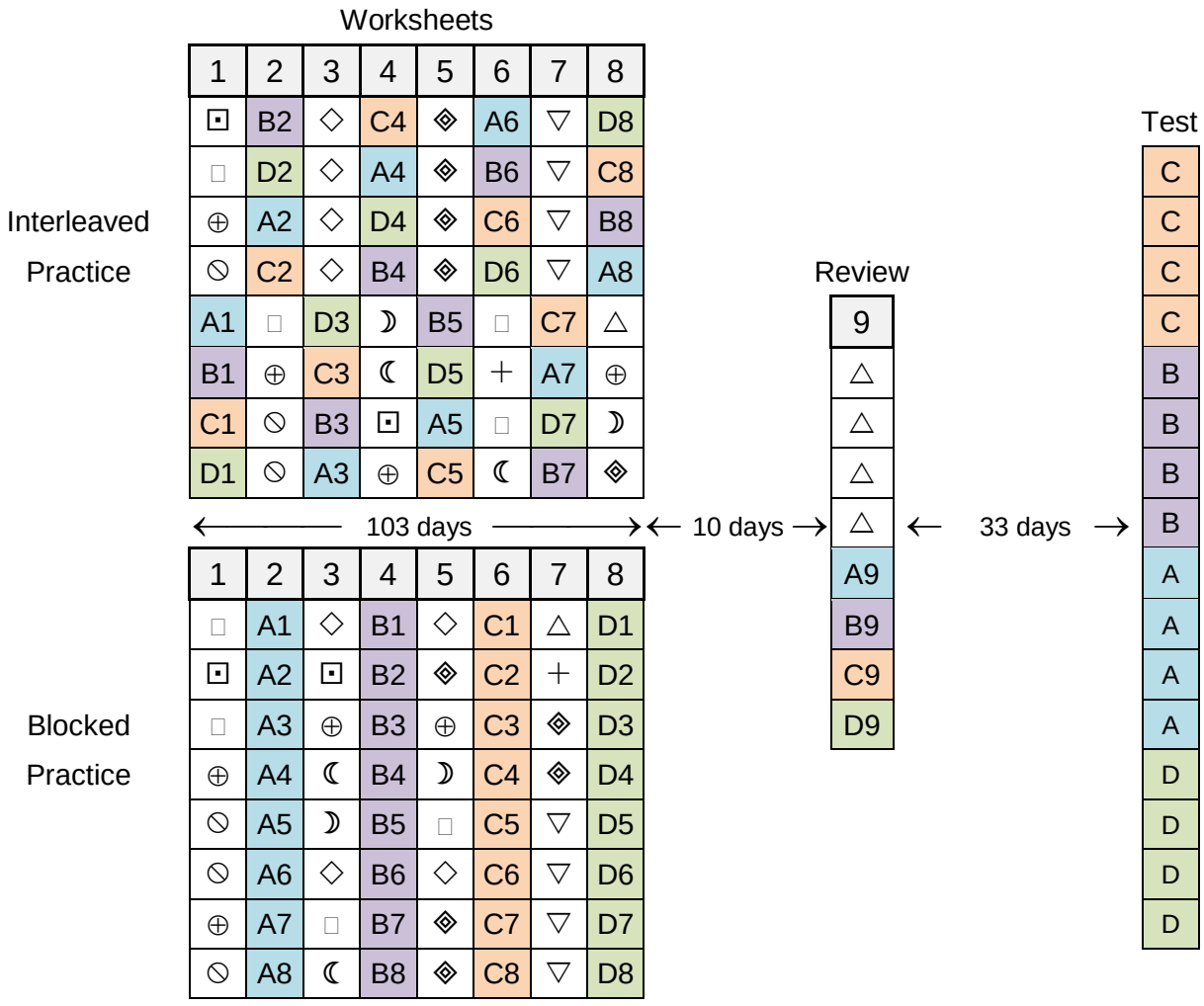



*Figure 2.* Procedure. Students completed a practice phase, review, and test. The duration of the time intervals shown above are means (see text). Each worksheet included 8 problems. The worksheets included critical problems, which were like the test problems, and filler problems. Each kind of critical problem is represented by letter A, B, C, or D (see Figure 1). For example, D5 represents a particular circle problem. Each kind of filler problem is represented by a unique symbol (e.g., the inverted triangle represents a probability problem).