4-2-2008

The Benefits of Interleaving Different Kinds of Mathematics Practice Problems

Kelli M. Taylor
University of South Florida

The Benefits of Interleaving Different Kinds of Mathematics Practice Problems

by

Kelli M. Taylor

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Doug Rohrer, Ph.D.
Michael Coovert, Ph.D.
James Eison, Ph.D.
Kristin Salomon, Ph.D.
Toru Shimizu, Ph.D.

Date of Approval:
April 2, 2008

Keywords: mixture, mixed practice, children, practice sets, textook design

Dedication

This dissertation is dedicated to my mom and dad who have always encouraged me to reach for the stars and have been there when I could not quite reach and needed a boost, to my grandfather who was always my biggest fan and whose faith in me never wavered, to my fiancé who never let me forget to enjoy myself, and to all my friends and family who have been the best cheerleaders in the world.

Acknowledgements

I would like to thank my committee members for all the helpful advice they have provided. I would also like to thank all the schools and camps who have given me the rewarding opportunity to work with their children. I would like to thank the U.S. Department of Education for sponsoring this research. Finally, I would like to thank my advisor for all of his guidance and patience over the last six years.

Table of Contents

List of Figures

The Benefits of Interleaving Different Kinds of Mathematics Practice Problems

Kelli Taylor

ABSTRACT

In most mathematics textbooks, virtually all of the problems in each set of practice problems, or in each *practice set*, relate to the immediately preceding lesson – an arrangement described here as the *standard format* of practice. Alternatively, the problems within a *shuffled* practice set are drawn from numerous lessons. With the shuffled format, each practice set has two distinguishing features: *within-session spacing*, in which problems *of the same kind* appearing in a single practice set are separated by some period of time, and *mixed practice*, in which different types of problems are interleaved. Although previous studies have assessed the combined effects of within-session spacing and mixed practice, the experiment presented here is the first to examine the effect of mixture while holding fixed the effect of within-session spacing in order to determine whether there is a benefit of mixture above and beyond the well-documented benefits of within-session spacing.

Fifth-grade students attended two sessions, a practice session and a test session, spaced one day apart. All students were taught how to solve four kinds of problems, and every student received the same tutorials and the same practice problems.

Students were randomly assigned to receive one of two kinds of practice: mixed practice, in which all four types of problems were interleaved, or unmixed practice, in

which all the problems of each type appeared in a block. Critically, in the unmixed practice condition, problems were separated by unrelated filler tasks so that the duration between each problem and the next problem *of the same kind* was equated for the mixed and unmixed conditions (i.e., the amount of spacing between two problems of the same kind was held constant). One day later, students returned for a test that included one novel problem of each kind, and, on average, the mixed practice group outscored the unmixed practice group by a large margin (77% vs. 38%). Thus, although there are limitations on the generalizability of the data, these findings nevertheless suggest that mixed practice, an important feature of shuffled practice sets, might boost mathematics proficiency.

Introduction

Investigations into the quality of mathematics education in the United States paint a distressing picture. Research shows students in the U.S. are not reaching the standards set by the U.S. Department of Education (e.g., Lee, Grigg, & Dion, 2007) and are performing below the standards of many other nations (e.g., Branigan, 2008). This poor performance of U.S. students has been documented with elementary school, middle school, and high school students (e.g., Lee et al., 2007; Gonzalez et al., 2004). For example the findings of a recent study conducted by the National Center for Educational Statistics revealed that approximately 60% of fourth-grade students and 68% of eighth-grade students in the United States are "not proficient" in the basic skills expected for their grade level (Lee et al., 2007). Similarly, in a study of 15-year-olds conducted by the Program for International Assessment, 26% of students in the U.S. failed to reach a level of proficiency described by the authors as basic (Gonzalez et al., 2004).

The poor condition of mathematics education in the United States becomes even more apparent when we compare the mathematics achievement of our students with that of students in other countries. The Third International Mathematics and Science Study (TIMSS) compared the mathematics achievement of fourth-grade students from 25 industrialized countries and eighth-grade students from 45 industrialized countries and the results were discouraging (Gonzalez et al., 2004). For instance, while 72% of students in Japan, the country with the highest mathematics literacy score, reached or excelled the intermediate benchmark in the study, only 60% of students in the U.S. reached this benchmark. Even more striking was the discrepancy between the percentage of Japanese (21%) and U.S. (7%) students who were "proficient" in the most advanced skills for their grade level (Gonzalez et al., 2004).

This lack of mathematics proficiency brings to the forefront the importance of improving mathematics education. This evidence raises the question, "What are these countries doing that we are not?" In an attempt to answer this question, eighth grade teachers from the countries that outscored the U.S. in the 1995 TIMSS study, which included Australia, the Czech Republic, Hong Kong SAR, Japan, the Netherlands, and Switzerland, were video taped while delivering one mathematics lesson, and a total of 638 mathematics lessons were analyzed (Silver, 2003). A major conclusion of this study was that there is no one method of instruction found in all six of the best-performing countries. In fact, there was not even consistency among the best-performing countries regarding many dimensions of instruction (Silver, 2003).

However, the instructional practices of the best-performing countries did share a few characteristics not seen in U.S. classrooms. For example, the amount of classroom time spent solving problems – either independently or in small groups – was greater in the best-performing countries than in the United States. This commonality notwithstanding, the results of the TIMSS study demonstrate that excellence can be achieved in many different ways.

One possible reason that the TIMSS video study did not reveal many commonalities among the best-performing countries is its singular focus on classroom activities without additional examinations of practice strategies. Indeed, the majority of research on mathematics education has focused on instruction rather than practice. Yet, mathematics students devote the majority of their effort to practice problems. Hence, it seems that many mathematics researchers arguably misplace the focus of their research, as there would appear to be much more to gain from assessing and improving the efficiency of mathematics practice.

Therefore, this study focused on mathematics practice, not instruction. Numerous features of mathematics practice can be manipulated, but this study focused on just one such feature: the ordering of practice problems in mathematics textbooks. More specifically, I explored whether interleaving many types of problems in a single practice set is more effective than not doing so.

Two Formats of Practice Sets

In virtually every mathematics textbook, each lesson is followed immediately by a set of practice problems, or *practice set*, and these practice sets are typically arranged in one of two fundamentally different ways. The more commonly used format for practice sets is dubbed here the s*tandard format*. In the standard format, practice sets contain problems drawn almost exclusively from the immediately preceding lesson. For example, in a statistics textbook, a practice set immediately following a lesson on the dependent samples t-test will consist almost entirely of problems related to the dependent samples t-test. Alternatively, and more rarely, the format of practice sets in some mathematics textbooks is described here as the *shuffled format*. Each shuffled practice set includes not only problems relating to the immediately preceding lesson but also problems relating to multiple previous lessons, and problems from these lessons are interleaved. For example, a shuffled practice set from a statistics textbook following the lesson on dependent samples t-tests would include no more than a few problems on the dependent samples t-test, and these problems would be mixed with many more problems drawn from previous lessons, such as problems on one-sample t-tests, dependent samples t-tests, confidence intervals, z-scores, and so forth, so that no two problems of the same type appear consecutively. Thus, any two problems on dependent samples t-tests would be separated by unrelated problems.

Importantly, a textbook using standard practice sets could be easily transformed into one using shuffled practiced sets. Doing so would not require changing the content (i.e., lessons and practice problems) of the textbook; instead, it would require only a reordering of practice problems. That is, a textbook with shuffled practice sets could include the exact same lessons in the same order and the exact same collection of practice problems as a textbook with standard practice sets, and the only difference between the two textbooks would be the ordering of these practice problems.

Three Critical Differences between Standard and Shuffled Practice Sets

Students completing a standard practice set rely on very different practice strategies than students completing a shuffled practice set. Therefore, in order to understand how the shuffled format can affect mathematics learning, it is important to first understand how each of these practice strategies affects mathematics learning. The sections immediately below detail the relevant literature for each practice strategy.

Overlearning

Compared to shuffled practice sets, standard practice sets include many problems of the same kind in consecutive order. For example, with standard practice, a lesson on the division of fractions might be followed by a practice set with dozens of consecutive fraction division problems. As an example, a practice set from a textbook in the popular series, *Everyday Math* (Dillard, 2007), includes 32 problems relating to the immediately preceding lesson. When students solve many problems of the same kind in immediate succession, they are relying on a practice strategy known as *overlearning,* which is the continued study or practice of some material *immediately* after students have achieved one success. Therefore, once students correctly solve the first problem, completion of every *immediate* subsequent problem *of the same type* constitutes overlearning.

By contrast, shuffled practice sets prevent students from using an overlearning strategy because a shuffled practice set includes more than a few problems from any particular lesson, and problems of any one kind are interspersed among other kinds. For example, with the shuffled format, a lesson on dividing fractions is followed by a practice set that includes no more than a few problems on the dividing of fractions intermixed with many more problems of different kinds, such as problems on adding fractions, multiplying decimals, measuring angles, and so forth.

Many educators and researchers advocate overlearning. For example, Fitts (1965) claimed that "The importance of continuing practice beyond the point in time where some…criterion is reached cannot be overemphasized" (p.195). Likewise, Foriska (1993) argued that overlearning is needed to move information from short-term to long-term memory. Finally, Hall (1989) claimed that overlearning can "….prevent significant losses in retention" (p.328). Many empirical studies in the verbal learning literature have resulted in a significant benefit of overlearning on retention. Although such studies do show that an overlearning strategy leads to better recall than lesser degrees of learning, these results are not surprising, as greater effort usually produces greater performance.

However, a careful review of the literature reveals that the apparent benefits of overlearning are subject to two critical caveats. First, the largest observed effects of overlearning tend to occur when students are tested after very short retention intervals. For example, 44 of the 51 studies in a meta-analysis of the overlearning literature by Driskell, Willis, and Copper (1992), relied on a retention interval of one week or shorter; moreover, the largest effects were observed after retention intervals of just seconds, minutes, or hours. Likewise, in Rohrer, Taylor, Pashler, Cepeda, and Wixted (2005), students who studied a list of 10 city-country pairs (e.g., *Doba*-Chad), 20 times (overlearners) scored higher than students who studied the list only 5 times (low learners) on a test one week later. However, studies that employ varying retention intervals show that the boosts in retention due to overlearning actually diminish with time (e.g., Craig, Sternthal, & Olshan, 1972; Reynolds & Glaser, 1964). For example, in the study reported above by Rohrer et al. (2005), the observed benefit of overlearning found at the one-week retention interval was not seen among students who were not tested until nine weeks after studying.

Another limitation of overlearning is that it is inefficient. That is, regardless of whether overlearning increases performance at a particular retention interval, the fact that overlearning requires additional study time demands that its benefits be weighed against its costs. Indeed, while many overlearning studies have shown that overlearning provides at least some benefit, the additional increase in retention is usually not proportionate to the required increase in study time. For example, in a study by Krueger (1929), when degree of overlearning was increased, retention also increased, but not proportionately to the increase in amount of overlearning. Specifically, when the number of trials increased by 50%, retention increased by 48%; yet, when the number of trials increased by another 33⅓%, retention increased by only an additional 27%. Many other studies have reported similar findings (e.g., Bromage & Mayer, 1986; Driskell et al., 1992; Kratochwill, Demuth, & Conzemius, 1977). In brief, while a little bit of overlearning might be advantageous, at least with brief retention intervals, extensive overlearning (i.e., heavy repetition) yields diminishing returns.

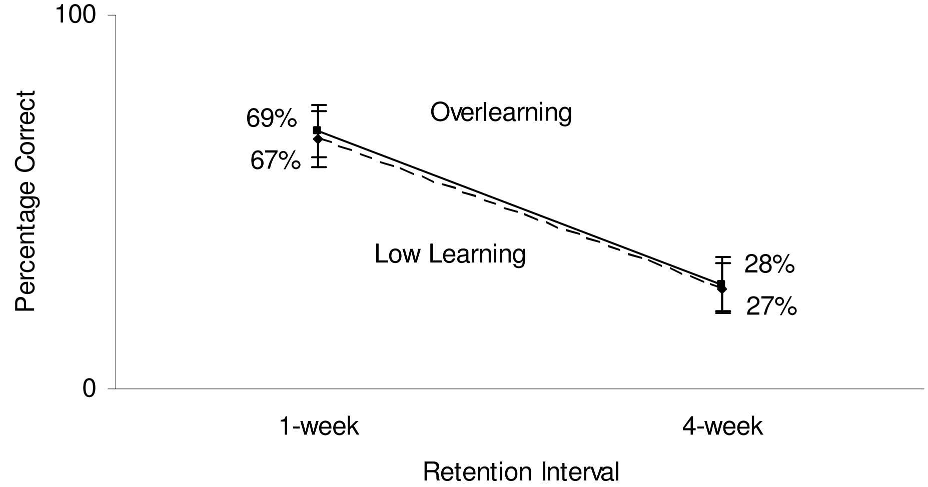
While the verbal learning literature suggests the benefits of overlearning are not commensurate with the increase in total study time required by overlearning, less is known about its benefits in the mathematics classroom because only two studies have examined the effects of heavy repetition on subsequent test performance. In a study by Rohrer and Taylor (2006), college students worked either three or nine practice problems of the same kind, in immediate succession, before returning for a test on a later date. The task required students to find the number of unique orderings of a sequence of letters with at least one repeated letter (e.g., the sequence abbccc has 60 permutations, including, abcbcc, acccbb, bbccca, and so forth). All students attended both a learning session and a test session. In the learning session, subjects first observed a tutorial before working either three problems (low learning) or nine problems (overlearning). Every student was given 45 s to work each problem, and, immediately after each problem, each student saw a visual presentation of the solution for 15 s. Students returned either one or four weeks later for a test (as determined by random assignment). The test consisted of five novel problems, and students received no feedback during the test. As shown in Figure 1A, completing an additional six practice problems provided no detectable benefit at either the one- or four-week retention interval. Thus, even after tripling the number of practice problems, heavy repetition – a key characteristic of standard practice – provided no observable benefit. Likewise, in two of the three conditions in an experiment reported by Rohrer and Taylor (2007), using the same permutation task described above, increasing the number of practice problems from two to four also provided no detectable benefit on a test given one week later (Figure 1B).

Thus, although the extant data are narrow in scope, the overlearning literature suggests that overlearning is an inefficient or even ineffective strategy for mathematics students. Yet, because overlearning is a defining feature of the textbooks relying on standard practice sets, it seems that a majority of mathematics students are wasting their time using an inefficient and ineffective practice method.

Within-Session Spacing

In standard practice sets, virtually every problem of the same kind appear consecutively, whereas, in shuffled practice sets, problems of the same kind are separated by at least one unrelated problem. Therefore, shuffled practice, but not standard practice, ensures that students rely on *within-session spacing*, whereby two problems of the same kind are separated by a task-filled delay. This practice strategy, within-session spacing, has been shown to benefit subsequent test performance for a wide variety of tasks and skills (e.g., Balota, Duchek, & Paulin, 1989; Glenberg, 1976; Hovland, 1940; Maskarinec 1985; Toppino, Kasserman, & Mracek, 1991). Within-session spacing has also been found to benefit both children (e.g., Rea & Modigliani, 1985; Seabrook, Brown, and Solity, 2005; Toppino, 1991; Toppino & DiGeorge, 1984) and adults (e.g., Glenberg, 1976; Maskarinec & Thompson, 1976).

A Mean Test Scores from Rohrer and Taylor (2006)



B Mean Test Scores from Rohrer and Taylor (2007)

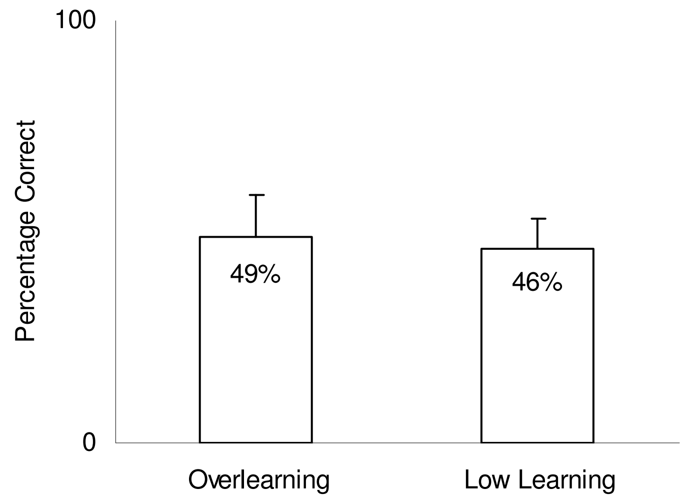


Figure 1. Previously published overlearning experiments. Vertical bars represent means. **A** Mean test scores from Rohrer and Taylor (2006). Performance did not differ reliably between the two groups at either retention interval. Error bars reflect plus or minus one standard error. **B** Mean test scores from Rohrer and Taylor (2007). Performance did not differ reliably between the two groups. Error bars reflect one standard error.

Like the literature on overlearning, the majority of findings reported in the literature on within-session spacing relate to tasks other than mathematics learning. In fact, only one known study examined mathematics learning. In a study reported by Rea and Modigliani (1985), third graders learned five multiplication facts (e.g., 8 x 5 = 40), and their practice of these facts was either massed or spaced within a single study session. In the spaced condition, there were either zero, one, two, or four distracter tasks (e.g., untested multiplication facts like 2x2), between two problems on the same multiplication fact, and in the massed condition, there were no distracter tasks between two problems on the same fact. On a test given one minute later, spaced practice led to better test scores than massed practice. However, as a caveat, the learning task used by these researchers is better described as a rote learning task rather than a mathematics task because subjects were asked to merely memorize and later recite five multiplication facts (e.g., 5 x 8 = 40). Nevertheless, given that within-session spacing benefits the learning of so many different kinds of skills, it seems reasonable to expect that the benefits of within-session spacing extend to mathematics learning as well. In summary, a review of the relevant literature shows that subsequent test performance benefits when the practice of some material is spaced throughout a session and not massed consecutively. Yet, while within-session spacing appears to be superior, standard practice sets rely instead on massing.

Mixed Practice

In a shuffled practice set, problems are drawn from multiple lessons and presented in an intermixed order. For example, in a statistics textbook using shuffled practice sets, the lesson on correlation might be followed by a practice set that includes many kinds of problems: correlation (cor), two-way ANOVA (2way), dependent samples t-tests (dep), independent samples t-tests (ind), and standard deviation (sd), and these different kinds of problems would be interleaved as follows:

cor, 2way, cor, dep, sd, dep, ind, sd, 2way, ind, dep, cor, ind, cor

By contrast, in standard practice sets, virtually all of the problems relate to the immediately preceding lesson; thus, a standard practice set following a lesson on correlation might look more like the following:

cor, cor, cor, cor, cor, cor, cor, cor, cor, cor, cor, cor, cor, cor

For example, in the fifth-grade workbook of the textbook series *Everyday Math* (Dillard, 2007), a 32-problem practice set follows the lesson on finding a fraction of a whole number (e.g., 2/3 of 20). Each of these problems involves this task; therefore, the practice set is in the standard format. The next practice set in the textbook consists of just six problems, five of which relate to other previous lessons, and one of which relates to the lesson on fractions making this a shuffled practice set. Thus, of the 38 problems following this unit, only six are in the shuffled format.

As a caveat, while standard practice sets do not often include problems on different topics, they do sometimes include different types of problems on the same topic. For example, many standard practice sets include a couple dozen problems on the same topic, which require the same procedure to solve, followed by a few word problems for that topic. For example, in the above mentioned standard practice set from the *Everyday Math* series (Dillard, 2007), the first 26 problems are simple fraction problems and the last six problems are word problems relating to the fraction task in the previous 26 problems; thus, in standard practice sets, even when there are different types of problems, all of the problems concern the same concept.

The vast majority of previous research on mixed practice examines the effects of this strategy on the learning of motor skills. For example, studies have found that mixed practice boosts performance on a bean bag tossing task (Carson & Wiegand, 1979), basketball shooting (Landin, Hebert, & Fairweather, 1993), and weapons firing (Keller, Li, Weiss, & Relyea, 2006).

Far fewer studies on mixed practice have examined the effects of this strategy on the learning of cognitive skills. For instance, van Morrienboer and Croock (1997) explored the effects of mixture on how well participants could troubleshoot a computer-based simulation and found a beneficial effect of mixture on the correct solution to questions involving the application of the learned skill to new situations.

Studies exploring the effects of mixed practice on mathematics learning are even rarer. In fact, there is only one known study has explored the effects of mixed practice on mathematics learning. In the study conducted by Rohrer and Taylor (2007), undergraduates first read a tutorial on how to find the area of four obscure solids: a wedge, a spherical cone, a spheroid, and a half cone. They then attended two practice sessions spaced one week apart and completed eight practice problems on each solid (a, b, c, d) in one of the two following orders:

	Session 1	Session 2
Mixed:	a b c d b d a c c a d b d c b a	a b c d b d a c c a d b d c b a
Unmixed:	a a a a b b b b c c c c d d d d	a a a a b b b b c c c c d d d d

Thus, the four kinds of practice problems were either blocked by kind or mixed together. As shown in Figure 2, students in the mixed practice condition outscored students in the unmixed practice condition on a test given one week later. Notably, this large boost in retention was achieved without having to solve additional problems.

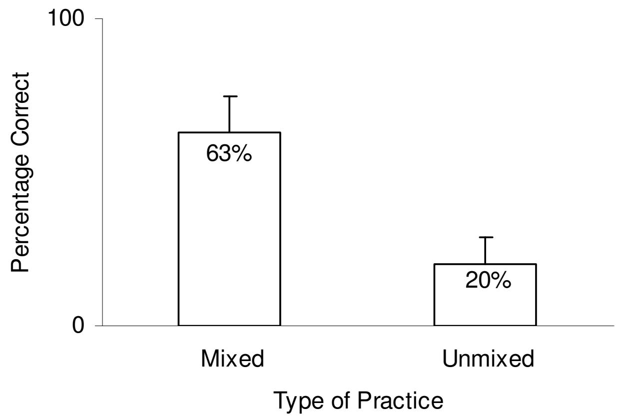


Figure 2. Previously published mixing experiment (Rohrer & Taylor, 2006). Vertical bars represent mean test scores. Test scores differed reliably between groups. Error bars reflect one standard error.

While previous studies have purportedly shown a benefit of mixed practice, each is confounded in a strict sense. Specifically, the design of these experiments ensured that the introduction of mixture also induced within-session spacing, which, as described above, has been shown to benefit subsequent test performance. For example, a group of practice problems from a statistics textbook are presented below in two orders, mixed and unmixed.

Mixed: cor, 2way, cor, dep, sd, dep, ind, sd, 2way, ind, dep, cor, ind, cor

Unmixed: cor, cor, cor, cor, 2way, 2way, sd, sd, dep, dep, dep, ind, ind, ind

With mixed practice, there is always a gap of at least one problem of a different topic separating any two problems on the same topic. For example, no two correlation problems (cor) appear consecutively in the mixed practice set shown above; thus, practice on correlation is spaced within-session. However, with unmixed practice, any two problems on the same topic appear consecutively; thus, there is no within-session spacing between correlation problems. To our knowledge, all of the studies on mixed practice confounded mixing and within-session spacing, and because within-session spacing reportedly benefits test performance of many tasks, the apparent benefit of mixture may be partly or even entirely due to the benefit of within-session spacing.

In summary, practice sets in the standard format promote three practice strategies shown to be inefficient: overlearning, massed practice, and unmixed practice. By contrast, practice sets in the shuffled format promote practice strategies shown to be very efficient: within-session spacing and mixed practice. While there is evidence suggesting that there are important combined beneficial effects of within-session spacing and mixed practice (Mayfield & Chase, 2002; Rohrer & Taylor, 2007), to my knowledge, no unconfounded empirical evidence exists that reveals a benefit of mixed practice independent of the effects of within-session spacing. Therefore, the current study focuses on assessing the effects of mixed practice while holding the amount of within-session spacing constant. In other words, the purpose of this study was to determine if mixing, per se, provides a benefit to mathematics learning above and beyond the benefit of the within-session spacing it intrinsically incorporates.

Method

Participants

Twenty-four fifth-grade students (12 boys and 12 girls) completed two sessions. Students' ages ranged from 10 to 11 years. An additional two students completed the first session but failed to return for the second session. All students received small gifts such as pencils and stickers for participating, but no rewards were based on the level of performance.

Task

Students were first taught the four formulas shown in Appendix A. Each formula is used to find the number of faces, corners, edges, or angles on a prism and includes the letter *b,* which stands for *base sides,* or the number of sides on the base of a prism. Appendix A provides an illustration, the definition, and the formula for each of these four features of a prism. Next, students were shown how to use each formula to solve problems like the one shown in the bottom portion of Appendix B. In these problems, the number of base sides was always between 3 and 30. As shown, each problem required a two-step solution: writing the formula and substituting the number of base sides for the letter *b* in the formula. In addition to the mathematics problems, students worked on various filler tasks, as explained in more detail further below (see Appendix C).

Procedure

Each student attended two sessions separated by one day: a *practice session* and a *test session*. Before the first session, each student was randomly assigned to either the mixed practice condition or the unmixed practice condition.

Test of Prior Knowledge. To verify that this task was unknown to our participant pool, a pre-test was given to all students at the beginning of the first session. Each student had two minutes to complete four problems – one problem of each kind. None of the students in this study answered any of these questions correctly. Therefore, it appears that this task was virtually unknown to the sample.

Practice Session. The practice session, which was devoted solely to the learning and practice of the task, included three phases: an *introduction phase*, a *formula-learning phase* and a *learning-to-substitute phase.* To ensure time on task was equal for students in both experimental groups, the computer paced all three phases of the practice session. During the introduction phase, every student observed a one-minute computer-paced visual presentation that included an introduction to prisms and the four features of a prism: faces, corners, edges, and angles. In both the formula-learning phase and the learning-to-substitute phase, the only difference between the mixed practice condition and the unmixed practice condition was the presentation order of these tutorials and problems.

The purpose of the formula-learning phase was to familiarize students with the four formulas they would be using during the learning-to-substitute phase. For students in both conditions, this phase included four 7-s tutorials during which the formula for each feature was presented (e.g., faces = b + 2), and thirty-two 10-s practice problems (i.e., eight problems on each of the four features). Appendix B shows an example formula-learning problem. During each of these practice problems, the name of a feature was displayed, and students had five seconds to write the correct formula for that feature (e.g., the word face appeared on the screen, and students were asked to write the right side of the correct formula, b+2), and five seconds to check their work. Students checked their work by writing a checkmark next to their answer if it was correct and an *x* next to their answer if it was incorrect. Furthermore, if their answer was incorrect, students wrote the correct solution in the space provided.

Students in the mixed practice condition viewed all four tutorials before beginning the 32 practice problems, which were in a mixed order. The orderings of the problems in the mixed and unmixed conditions appear below. The capital letters A, B, C, and D represent the tutorials for each of the four different features, the lowercase letters a, b, c, and d represent problems on the four different features, and each letter f represents a 30-second filler task.

Mixed: ABCDabcdcadbdcbabdacbdacdcbacadbabcdffffffffffffffffffffffffffff

Unmixed: AafafafafafafafaBbfbfbfbfbfbfbCcfcfcfcfcfcfcDdfdfdfdfdfdfd

The ordering of the problems had two characteristics. First, each set of four consecutive problems (e.g., 1 - 4, 5 - 8, etc.), included one problem of each kind (i.e., one face, one corner, one edge, and one angle). Second, the order of these problems ensured that there were three problems, on average, between every two problems of the same kind. For example, as shown above, the eight corner problems (indicated by a lowercase c) were separated by gaps of 1, 4, 5, 3, 1, 2, and 5 intervening problems, respectively, and the average of these gaps was three intervening problems, or 30 seconds. After completing all 32 formula-learning problems, students in the mixed condition completed 14 minutes of filler tasks.

By contrast, as shown above, students in the unmixed condition practiced recalling the formulas in a blocked order. They viewed the tutorial for the first feature immediately before completing the eight practice problems for that feature. Next, they viewed the tutorial for the second feature and then immediately completed the eight practice problems for that feature. This process continued for all four features. Importantly, 30 seconds of filler tasks separated each formula-learning problem on the same feature. This guaranteed that the interval between two problems of the same kind was, on average, equivalent in both the mixed and unmixed conditions. In other words, within-session spacing was held constant.

The learning-to-substitute phase began immediately after the formula-learning phase ended. During this phase, students viewed a tutorial on how to substitute the number of base sides for the letter *b* in each formula. As in the formula-learning phase, the tutorials, the problems, and the amount of within-session spacing were the same in both conditions. All students viewed four 30s tutorials – one tutorial explaining how to substitute for *b* in the formulas for each of the four features – and all students completed twelve 25-second *substitution* problems – three problems of each kind. Twenty-five seconds was allotted for each problem during the learning-to-substitute phase. Students had 15s to complete each problem, and, immediately afterwards, students saw the correct solution and then, if necessary, corrected their work (10 s). A sample problem is shown in Appendix B. As in the formula-learning phase, the only difference between the two conditions was the order in which the tutorials and problems were completed. Again, students in the mixed practice condition viewed all four tutorials before beginning the 12 practice problems. The practice problems were ordered with the same constraints of the formula-learning phase so that in every block of four problems, there was one problem on each feature, no two problems of the same kind appeared consecutively, and, on average, there were three problems, or 75 seconds, separating two problems of the same kind. After completing all 12 substitution problems, students in the mixed condition then completed 10 minutes of filler tasks. The orderings of the problems in the mixed and unmixed conditions appear below. The letters A, B, C, and D represent the tutorials for each of the four different features, the lowercase letters a, b, c, and d represent problems on the four different features, and the letter f represents a 75-second filler task.

Mixed: ABCDabcdbdacabcdffffffff

Unmixed: AafafaBbfbfbCcfcfcDdfdfd

By contrast, students in the unmixed condition completed their problems in a blocked order. As shown above, they viewed the tutorial on each feature immediately before beginning the three practice problems on that feature. Again, for both conditions, the average interval between two problems of the same kind was equal. Thus, during the learning-to-substitute phase, students in the unmixed condition completed 75 seconds of filler tasks between each problem of the same kind.

Test Session. One day after the practice session, students took three tests. None of the problems from the practice session appeared on any of the tests; therefore, only novel problems appeared on each test. Feedback was not provided during any of the tests.

Test 1, the *primary test*, required students to solve four novel substitution problems (one face, one corner, one edge, and one angle) identical to those in the learning-to-substitute phase. That is, for each problem, students were required to recall the correct formula and then substitute the correct number of base sides for the letter *b* in the formula. Two minutes were allotted for completion of this test. Because the only difference between the two experimental conditions was the ordering of the problems, a reliable difference between groups on this test would allow us to attribute the effect to the mixed ordering of the problems during practice, and not to the amount of within-session spacing between each problem of the same kind.

On Test 2, the *substitution only test*, the formula needed to solve each problem was provided at the top of the page (e.g., faces=b+2). Thus, unlike the primary test, the substitution only test did not require students to recall the correct formula. Therefore, to complete the substitution only test, students needed to know *how* to use each formula and not *which* formula was appropriate. Thus, a reliable difference between students in the mixed and unmixed conditions on the primary test but not on the substitution only test suggests the benefits of mixed practice arise from the discrimination training provided by mixed practice but not unmixed practice. That is, in the mixed practice condition, two problems of the same kind never appeared consecutively, so students needed to rely on the key elements of each problem to choose the correct formula, but, in the unmixed condition, all of the problems of one kind appeared in succession, and rather than using the details of the question to choose the correct formula, students merely repeated the use of the same formula for every problem in each block of problems on the same feature

Test 3, the *general mathematics test,* assessed the students' general mathematics ability. The purpose of this test was to uncover any differences in mathematical ability of the two groups (which would presumably be small because of random assignment). The test consisted of 10 questions drawn from the 2005 and 2007 National Assessment of Educational Progress, or the *NAEP*, (National Center for Education Statistics, 2007). These questions for Test 3 were chosen from the three sections of the NAEP that were most similar to the prism task: number properties and operations, algebra, and geometry.

Results

Practice Performance

There were eight formula-learning problems (F1-F8 in Figure 3) and three learning-to-substitute problems (S1-S3 in Figure 3) for each of the four kinds of problems (i.e., face, corner, edge, and angle), and the mean accuracy for these 11 problems, averaged across the four kinds, is plotted in Figure 3. For the eight problems in the formula-learning phase, mean accuracy for the mixed practice condition (68%) was reliably lower than that in the unmixed practice condition (99%), $t(22) = 4.94$, $p < .01$, Cohen's $d = 2.02$. Likewise, during the learning-to-substitute phase, the mean accuracy for students in the mixed practice condition (79%) was reliably lower than that of the students in the unmixed practice condition (98%), $t(22) = 2.94$, $p < .05$, Cohen's $d= 1.20$. Finally, mean accuracy on the third and final round of learning-to-substitute problems (i.e., the final face, edge, corner, and angle learning-to-substitute problems), was reliably lower for the mixed practice group than for the unmixed practice group , $t(22) = 2.46$, $p < .05$, Cohen's $d = 1.01$. In summary, mixture *impaired* performance during the *practice* phase.

Test Performance

By contrast, mean performance on Test 1 (the primary test), which included items just like those in the second phase of the practice session (learning-to-substitute phase), was greater for students who relied on mixed practice rather than unmixed practice. In fact, mixed practice proved superior by an extremely large margin (77% vs. 38%), $F(1, 21) = 7.43$, $p = .013$, Cohen's $d = 1.11$ (Figure 3). In order to ensure that this difference was not due to pre-experimental differences in ability, these data were analyzed using a one factor fixed effect analysis of covariance (ANCOVA) with general mathematics test score as the covariate. There was no significant effect of the covariate, as the general mathematics test did not provide significant regression effects for scores on the primary test.

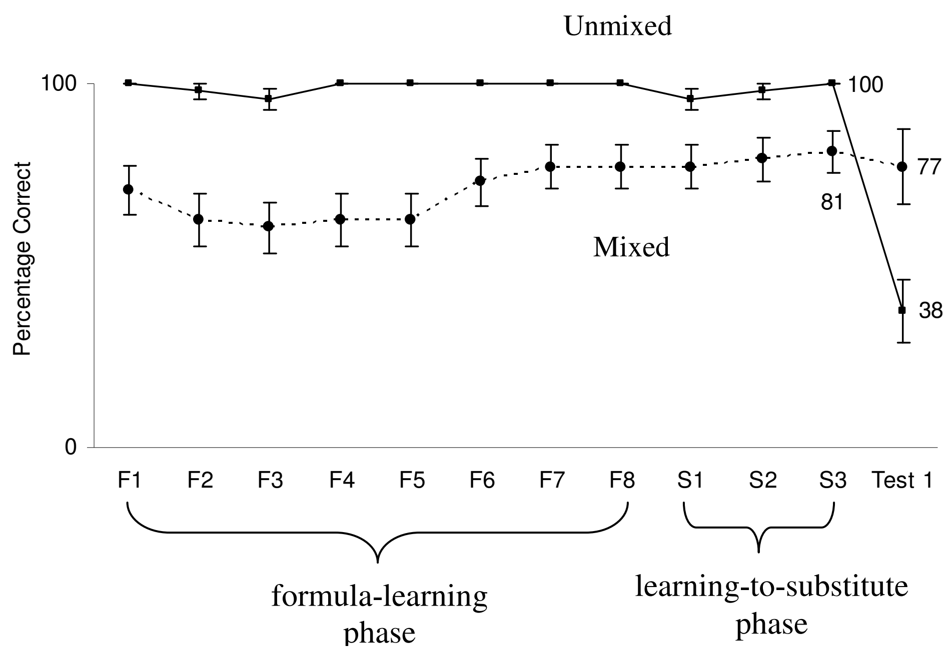


Figure 3. Practice and test results. F1 through F8 represent the mean scores for the eight formula-learning problems of each kind (face, corner, edge, angle), with data averaged across the four tasks. Likewise, S1 through S3 correspond to the three learning-to-substitute problems for each of the four tasks. Error bars equal plus or minus one standard error.

Except for a single problem left blank by a single student, every error on Test 1 was one of two kinds. With the first type of error, or a *task-related formula error,* students mistakenly chose the formula for one of the other three tasks they had learned - for example, choosing the "face formula" for a "corner problem." With the second type of error, or a *task-unrelated formula error*, students used a formula other than one of the four used in this study. Figure 4 shows the mean percentage of test problems answered incorrectly due to each of these types of errors. This percentage differed reliably between the mixed and unmixed conditions, $t(22) = 4.20$, $p<.01$, Cohen's $d = 6.41$.

As shown in Figure 4, students in the unmixed condition answered 46% of their questions incorrectly due to a task-related formula error, whereas this percentage was much smaller in the mixed condition (10%), $t(22) = 3.59$, $p<.01$, Cohen's $d = 1.46$. In other words, the unmixed group was much more likely than the mixed group to use a formula from the learning phase of this study for the wrong type of problem.

Moreover, there was no reliable difference in percentage of task-unrelated formula errors between the two groups. That is, students in the mixed and unmixed conditions committed about same percentage of task-unrelated formula errors. Therefore, these results are consistent with the hypothesis that the benefits of mixed practice stem from the training it provides on discriminating between superficially similar formulas.

By contrast, there were no reliable differences between the mixed and unmixed practice groups on the substitution only test or on the general mathematics test. As in the practice session, no statistically significant gender differences were observed for any of the three tests.

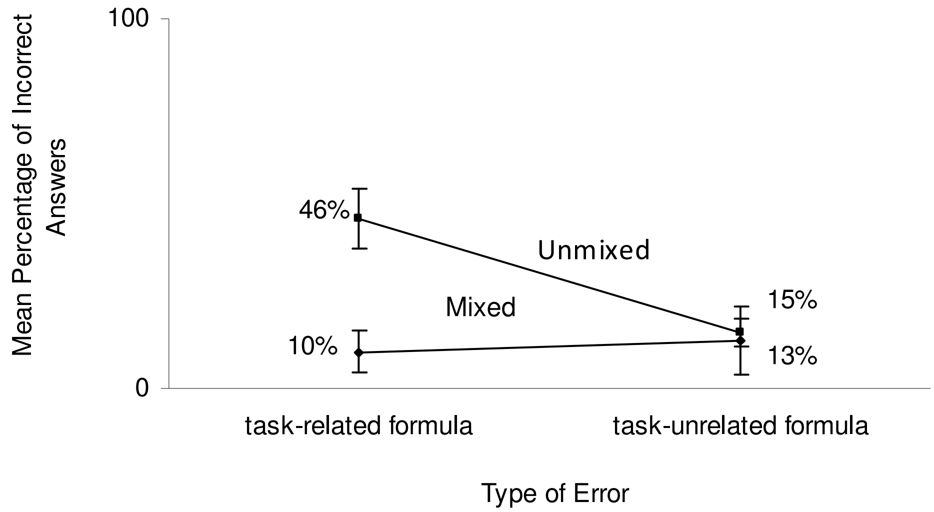


Figure 4. Analysis of test errors. Data points represent the percentage of problems on each test that were answered incorrectly due to task-related or task-unrelated formula errors as average across all participants in an experimental group. Error bars equal plus or minus one standard error.

Discussion

Summary of Results

While mixed practice impeded *practice* performance, the scores on the critical subsequent test (Test 1) were much greater when practice was mixed rather than unmixed. This benefit of mixed practice can be attributed to the mixed ordering of the problems in the mixed condition and not to the degree of within-session spacing, which is inherently introduced when practice is mixed, because within-session spacing was held constant across both experimental groups. Therefore, unlike previous experiments that confounded mixed practice and within-session spacing, the current study demonstrates the independent effects of a mixed ordering of problems.

Moreover, because the primary test (Test 1), but not the substitution-only test (Test 2), required students to pair the appropriate formula with the appropriate kind of problem, the benefits of mixing on Test 1 but not Test 2 are consistent with the idea that mixing improves discriminability. That is, because the mixed practice group outperformed the unmixed practice group on the test that required students to discriminate between the different kinds of problems (Test 1) but not the test that required every necessary skill other than discrimination (Test 2), the benefit of mixed practice found here suggests that mixed practice improves learning by improving students' ability to discriminate between different kinds of problems. Thus, by this account, mixing improves one key aspect of learning but has no effect on retention.

In mathematics, the ability to discriminate between superficially similar problems is important because mathematics is characterized by superficially similar kinds of problems that require different kinds of solutions. For example, in an introductory statistics course, students learn about three kinds of t-tests: one-sample, independent samples, and dependent samples. While each of these problems requires different procedures and formulas to solve, they can all look very similar. For example, all three kinds of problems, as well as other problems requiring null hypothesis testing, usually include a data set, and an instruction to determine whether the data are statistically significant. Thus, because the statement of these problems does not indicate which statistical test is appropriate, students must rely on the characteristics of the problems to choose the appropriate solution. Though critically important, discrimination training is not provided by textbooks that use the standard format of practice sets (which are unmixed). For example, a practice set following the lesson on the independent samples t-test would include only independent samples t-test problems. Therefore, students know in advance which statistical test is needed for each problem and, as a result, do not need to understand certain key details of the problem. Thus, with the most commonly used standard format of practice sets, students receive practice on *how* to solve each problem but not on choosing *which* statistical test is appropriate.

By contrast, if textbooks adopted the shuffled format of practice sets, they would employ mixed practice and, as a result, discrimination training. This discrimination training is inherently part of a shuffled format because mixed practice sets include problems drawn from many different lessons. For example, a mixed practice set following the lesson on independent samples t-tests would likely also include problems relating to one-sample and dependent samples t-tests, and these three kinds of problems would be interleaved within a single practice set. Therefore, unlike with a standard practice set, when completing a shuffled practice set, students do not know in advance which statistical procedure is appropriate for each problem and must instead rely on the nature of each problem to choose the appropriate statistical test. In other words, when problems of different kinds are intermixed, each problem requires students to decide *which* procedure is appropriate as well as *how* to perform the procedure.

VandoerStoep and Seifert (1993) demonstrated the importance of this ability to discriminate between superficially similar problems in an experiment in which students learned to solve either a pair of superficially similar or superficially different mathematics problems by viewing either a tutorial emphasizing *how* to solve each kind of problem or a tutorial emphasizing *which* procedure was appropriate to solve each problem. When the two problems were superficially similar, the tutorial that emphasized which procedure was appropriate led to higher test scores than the tutorial emphasizing how to solve each problem, but when the two problems were superficially different, there was no reliable difference in effectiveness of the two tutorials. Therefore, when problems are superficially similar, as are many problems in mathematics textbooks, the discrimination training provided by mixed practice is useful in boosting subsequent test performance, as witnessed in the current study.

Costs and Benefits of Mixed Practice on Educational Outcomes

When assessing the utility of certain practice strategies, researchers often focus solely on effectiveness and overlook the importance of efficiency. However, because time is often a limiting factor in education, efficiency is an important characteristic of any practice strategy. If a practice strategy produces gains in learning or retention, but the magnitude of these gains do not meet or exceed the increase in time or effort required by the strategy, then the strategy is inefficient. For example, as discussed in the introduction, investigations of the overlearning strategy often report increases in retention of material, but these increases are rarely ever proportionate to the increase in study time required by an overlearning strategy (e.g., see Bromage & Mayer, 1986; Driskell et al., 1992; Kratochwill et al., 1977; Krueger, 1929).

The results of the current study suggest that mixed practice boosts efficiency because mixed practice requires the same number of problems and the same amount of study time as unmixed practice. Therefore, simply reordering the problems used in the unmixed condition caused test performance of students in the mixed condition to nearly double. The large benefit of mixed practice documented here is even more impressive when you consider that time was added between each problem of the same kind in the unmixed condition in the form of filler tasks to equate the within-session spacing between the two groups.

Yet another prerequisite of mixed practice is that it may reduce boredom. Unmixed practice sets are repetitive and perhaps monotonous. They often include dozens of problems requiring the same steps. By contrast, mixed practice sets consist of numerous different kinds of problems, and these problems are intermixed. Therefore, with a mixed practice set, students constantly have to switch gears and use different skills. For example, when completing a mixed practice set in statistics, students may move from work on z-scores to ANOVA, to independent samples t-tests, to correlation, to dependent samples t-test. This variety found in mixed practice sets might hold a student's attention longer than the monotony found in unmixed practice sets.

On the other hand, mixed practice is arguably more demanding than unmixed practice. In this study, in fact, accuracy on the *practice problems* was worse for the mixed practice group. Others have observed the same pattern – mixture impairs practice performance but boosts test performance – which is a finding described by Bjork and his colleagues as a "desirable difficulty" (e.g., Bjork, 1994; Christina & Bjork, 1991)

Another advantage of mixed practice sets is that they allow word problems to serve their appropriate purpose - promoting transfer of basic mathematical skills to more complex contexts. Unmixed practice sets usually include word problems requiring application of a skill to a more complex problem, but these word problems usually follow dozens of other problems requiring that skill. For example, in a practice set drawn from one of the *Everyday Math* textbooks (Dillard, 2007, p. 190) the first 26 problems ask students to do the same thing, find the value of a certain fraction of a whole number. The next six problems are word problems that require students to apply this skill to real-life scenarios; however, having completed 26 previous problems requiring this same procedure, students likely need not even read the word problem to identify the appropriate procedure needed to solve the problem. By contrast, in a mixed practice set, word problems would be intermingled in a practice set containing problems from many topics. Therefore, students must read and evaluate each word problem before deciding which procedure to apply to find the solution. Hence, each word problem in a mixed practice set is serving its purpose, teaching students how to apply basic mathematics skills to novel situations.

While there seem to be many benefits of mixed practice, there might be costs as well. For instance, as found in the present study (as well as previous studies, e.g., Schmidt & Bjork, 1992), mixing practice slows acquisition and therefore delays mastery. Furthermore, the introduction of mixture would presumably increase the time and effort needed to complete a practice set.

An Important Feature of Textbooks Using Shuffled Practice Sets

When the shuffled format of practice sets is applied to an entire textbook, students are forced to rely on another well-known and efficient practice strategy, *across-session spacing*, or the distribution of the practice of material across multiple practice sessions. Unlike textbooks using standard practice sets, in which all of problems of one kind in the textbook are usually massed into the same practice set, shuffled practice sets ensure that problems of the same kind are distributed across multiple practice sets because only a few problems on Topic A appear in the practice set immediately following the lesson on Topic A, and the rest of the problems on that topic appear in multiple subsequent practice sets distributed throughout the textbook. For example, after a lesson on the factoring of polynomials within a textbook using shuffled practice sets, only a few problems on polynomical factoring will appear in the practice set immediately following the lesson, and the remaining 25 or so problems on this topic will be dispersed among 10 or 20 subsequent practice sets; thus, practice on this skill is spaced across many practice opportunities.

As a caveat, many standard practice sets do include occasional sets of review problems, but review problems for each kind of problem usually appear just once (e.g., in the chapter review); moreover, the review problems constitute only a small percentage of the problems in the text. Thus, textbooks using shuffled practice sets allow for much more across-session spacing than textbooks using standard practice sets.

Numerous experiments have shown that spacing a given amount of practice across multiple sessions instead of massing this practice into a single session produces superior performance on a subsequent test– an effect known as the *spacing effect* (see reviews in Bjork & Allen, 1970; Cepeda et al., 2006; Dempster, 1987, 1988; Donovan & Radosevich, 1999; Mumford, Costanza, & Baughman, 1994; Raajimakers, 2003). This phenomenon has been studied in many areas of learning, such as verbal learning (e.g., Bahrick, Bahrick, Bahrick, & Bahrick et al., 1993; Bahrick & Phelps, 1987; Greene, 1989), motor tasks (e.g., Lee & Genovese, 1988), puzzle solving (e.g., Cook, 1934), science education (e.g., Reynolds & Glaser, 1964), and many others. In fact, the spacing effect has been observed in so many different types of tasks that some have called it robust and "ubiquitous in scope" (Dempster & Farris, 1990, p. 97). Because spacing provides multiple review opportunities on the same topic and has been reported to boost retention on so many diverse tasks, it is no wonder that it is championed as a great study method in the learning literature (e.g., Bahrick et al., 1993; Bjork, 1979, 1994; Bloom & Shuell, 1981; Dempster, 1988; Schmidt & Bjork, 1992).

There is also empirical evidence showing a benefit of across-session spacing on the retention of mathematics skills (e.g., Rohrer & Taylor, 2006; Rohrer & Taylor, 2007). In the study reported by Rohrer and Taylor (2006), for instance, 116 college students were randomly assigned to one of two groups that worked 10 problems in a single session (massed practice) or divided these 10 problems across two sessions separated by one week (spaced practice). The task required students to find the number of unique orderings of a sequence of letters with at least one repeated letter (e.g., the sequence abbccc has 60 permutations, including, abcbcc, acccbb, bbccca, and so forth). All students attended three sessions: two practice sessions and one test session. One week separated the two practice sessions, and students were randomly assigned to return either one or four weeks after the final practice session for a test. During both practice sessions, students had 45 s to solve each practice problem and 15 s to review the correct solution to that problem. The test included five novel problems without feedback. For students tested one week after the final practice session, there was no apparent benefit of spaced practice. However, for students tested four weeks later, spaced practice produced a large benefit for test performance.

This increase in the size of the spacing effect with longer retention intervals is commonly found in the spacing literature (e.g., Dempster, 1988), and it demonstrates that spacing improves long-term retention. Thus, the use of spacing would improve student performance on standardized exams, which often require students to remember information from previous years and not just the last few weeks. The results of Rohrer and Taylor (2006) suggest that across-session spacing, which would occur in a textbook with shuffled practice sets, is a highly efficient practice strategy, not only for verbal learning tasks, but also for mathematics tasks.

Adopting Shuffled Practice Sets

Review of past literature and the results of the current study suggest that students' mathematics proficiency could be improved if shuffled practice sets were incorporated into mathematics textbooks, and doing so can be easily accomplished by transforming current editions of mathematics textbooks. This transformation would require only a reordering of the extant practice problems in a textbook because the lessons and the collection of practice problems could remain the same. Textbook creators could use the following simple formulas to choose the problems that make up each practice set. In the example below, “Lesson N” represents the current lesson, and each practice set would include the following problems:

Six from Lesson N

Three each from Lessons N-1, N-2, and N-3

Two each from Lessons N-5, N-6, N-7, N-8, and N-9

One each from Lessons N-10, N-15, N-20, N-25, and N-30

Furthermore, when constructing practice sets for lessons early in the textbook, publishers could include problems that review concepts learned in previous years.

Successful Use of Shuffled Practice Sets in Textbooks

While I know of no textbook that utilizes shuffled practice sets exactly as laid out above, two textbook series do incorporate some of the features of shuffled practice. First, the Saxon series of mathematics textbooks incorporates across-session spacing (Saxon, 1997). For example, in the Saxon Algebra I textbook (Saxon, 1997), the practice set following Lesson 66 includes only three problems related to Lesson 66, and the remaining 15 problems are drawn from many previous lessons and constitute across-session spacing. Secondly, the majority of problems on any given topic are interleaved with problems on other topics in Saxon textbooks. That is, when multiple problems from a given lesson appear in a Saxon practice set, these problems are blocked together. However, only a small percentage of the total number of problems in the textbook from a given lesson appear in practice sets together, and the rest of the problems from that lesson appear as the only problem from that lesson in a practice set.

Results of studies on the effectiveness of the Saxon series (1997) have produced mixed results. For example, according to a report from the U.S. Department of Education that summarized dozens of previous studies comparing a Saxon textbook and one utilizing standard practice, Saxon proved superior in some of the studies with middle school students but not in the studies with elementary school students (Saxon Elementary School Math, 2007; Saxon Middle School Math, 2007). While informative, such studies comparing Saxon and standard textbooks do not speak to the benefit of shuffled practice because any two textbooks differ in many ways, such as teacher instruction, textbook lessons, the choice of practice problems, and so on. Therefore, such studies do not assess the relative merits of shuffled and standard practice .

Another textbook series that utilizes some features of shuffled practice is *Everyday Math* (Dillard, 2007). This series includes many mixed practice sets which include approximately six problems drawn from six previous lessons, although the practice sets that immediately follow each lesson are standard practice sets, which include only problems from the immediately preceding lesson. At least one study has found an advantage of the Everyday Math textbook over another textbook (Riordan & Noyce, 2001), but, as noted above, such a study does not explicitly compare standard and shuffled practice.

Limitations

There are several limitations to the generalizability of these data. First, this experiment did not directly compare a truly unmixed and mixed practice set as found in textbooks. That is, the unmixed practice sets used in the current experiment include within-session spacing, which is not included in unmixed practice sets found in textbooks. This characteristic of the unmixed practice was added to avoid confounding practice method and within-session spacing. However, because studies have shown that within-session spacing can boost mathematics performance (Rea & Modigliani, 1985), and mixed practice inherently incorporates within-session spacing, it is likely that the magnitude of the effect would be greater in a direct comparison of these two types of practice sets. Another limitation on the generalizability of the present results is that it remains unknown whether the benefits of mixture would extend to students of a different age or race. That is, the sample included only 10- and 11- year-old fifth graders at a private school, and all but one of these participants were White. Finally, although this study was conducted in a classroom, it was, for all intents and purposes, a laboratory study. For example, to ensure control of many variables, the lesson was delivered by a computer, not an instructor. Furthermore, students had only a single practice session to practice the material, whereas a shuffled textbook would provide multiple practice opportunities. Finally, the rate of practice was paced by the computer, so students could not work at there own pace, as is usually the case when students work on homework assignments.

Finally, the task used in the current study was a procedural task rather than a conceptual task. As defined by Rittle-Johnson, Siegler, and Alibali (2001), procedural knowledge is “the ability to execute action sequences to solve problems,” whereas conceptual knowledge requires “implicit or explicit understanding of the principles that govern a domain and of the interrelations between pieces of knowledge in a domain” (pp.346-347). They claim procedural knowledge is not widely generalizable because it is "tied to specific problem types" (p. 346) whereas conceptual knowledge is more generalizable because it is not.

These two types of knowledge – procedural and conceptual – are often viewed as a dichotomy in which conceptual knowledge is preferred over procedural knowledge (e.g., Xin, 2007). However, many researchers disagree with this dichotomous classification of procedural and conceptual knowledge (e.g., Rittle-Johnson, Siegler, & Alibali, 2001; Rittle-Johnson & Star, 2007; Wu, 1999). Wu (1999) claims that students need to master basic skills, or procedural knowledge, before they can acquire a deeper understanding of underlying mathematics concepts.

Similarly, Rittle-Johnson et al. (2001) do not view procedural and conceptual knowledge as mutually exclusive. Instead, they view these two types of knowledge as two strongly connected points on opposite ends of a continuum and suggest that there is an iterative relationship between procedural and conceptual knowledge. In their words, "Conceptual and procedural knowledge develop iteratively, with increases in one type of knowledge leading to increases in the other type of knowledge, which trigger new increases in the first" (Rittle-Johnson et al., 2001, p.346). With this model, both types of knowledge are very important for mathematics education. Students must have a strong procedural knowledge of a concept or skill before they can develop higher-order understanding of that concept or skill, and students must have conceptual knowledge of a skill before they can successfully use their knowledge to apply that skill to more difficult applications. Thus, educators should not discard the importance of the facilitation of *both* types of knowledge.

While the current study does not assess the effects of mixed practice on the retention of conceptual knowledge of a skill, educators should not overlook the findings reported here. As evidenced by scores on the primary test, students in the mixed practice condition were better able to use the procedure required by the prism task (i.e., choosing the correct formula and substituting for *b*), than students in the unmixed condition. The iterative model suggested by Rittle-Johnson et al. (2001) suggests students must form basic procedural knowledge of a skill or concept before forming conceptual knowledge of that skill or concept. Therefore, following mixed practice, students have a stronger understanding of the procedural requirements of the task and are thus better prepared to begin building a conceptual understanding of the task. That is, rather than focusing energy on remembering that the formula for finding the number of faces on prism is faces = b + 2, students can now focus energy on understanding why this is the formula.

Future Directions

Therefore, because it remains unknown whether the benefits of shuffled practice would replicate under more general circumstances, further studies must examine whether the present results generalize to broader samples, longitudinal studies, and more complex tasks. First, because this study focused on children at just one grade level, it is not known if the effects of mixed practice are qualitatively different for students of different ages. Therefore, the effects of mixed practice should be explored at many grade levels from early elementary school to high school. Second, when the shuffled format of practice is used, students practice each skill in multiple practice sets. In the current study, students attended only a single practice session. Therefore, we do not know how the use of this across-session spacing factor of the shuffled format will affect subsequent test performance. Therefore, a study is needed in which practice problems on multiple skills are spaced across many practice sessions, possibly across a semester. Third, as mentioned above, the task used in the current study was procedural, so it remains to be seen if the benefits of mixed practice would generalize to conceptual tasks. Therefore, future studies should include more complex tasks involving conceptual learning.

Brief Summary

The shuffled format of mathematics practice sets ensures that, students rely on three practice strategies: mixed practice, within-session spacing, and across-session spacing. While previous studies have provided empirical evidence supporting the use of across- and within-session spacing, the present study is the first to demonstrate a benefit of mixed practice when the degree of within-session spacing is held constant. Based on these findings, shuffled practice deserves greater attention from mathematics researchers, educators, and textbook publishers.

References

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

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

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology & Aging, 4,* 3-9.

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

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

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*, 567-572.

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

Branigan, C. (Ed.). (2008). Videos show best math teaching practices from other countries. *eSchool News*, Retrieved January 8, 2008, from http://www.eschoolnews.com.

Bromage, B. K., & Mayer, R. E. (1986). Quantitative and qualitative effects of repetition on learning from technical texts. *Journal of Educational Psychology, 78,* 271-278.

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

Cepeda, N. J., 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.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23-56). Washington DC: National Academy Press.

Cook, T. W. (1934). Massed and distributed practice in puzzle solving. *Psychological Review, 41*, 330-355.

Craig, C. S., Sternthal, B., & Olshan, K. (1972). The effect of overlearning on retention. *Journal of General Psychology, 87,* 85-94.

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

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

Dempster, F. N., & Farris, R. (1990). The spacing effect: Research and practice. *Journal of Research and Development in Education, 23,* 97-101.

Dillard, A. et al. (2007). *Student Math Journal: Everyday Math: The University of Chicago School Mathematics Project, Third Edition*. Chicago: Wright Group/McGraw-Hill. Copyright 2007 by Wright Group/McGraw-Hill.

Dovovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect. *Journal of Applied Psychology, 84*, 795-805.

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

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

Foriska, T.J. (1993). What every educator should know about learning. *Schools in the Middle, 3,* 39-44.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15,* 1-16.

Greene, R. L. (1989). Spacing effects in memory: evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 371- 377.

Gonzalez, P., Guzman, J. C., Partelow, L., Pahike, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights from the trends in international mathematics and science study: TIMSS 2003*. Retrieved December, 12, 2007 from http://nces.ed.gov/surveys/pisa/pdf/comppaper12082004.pdf.

Hall, J. F. (1989). *Learning and Memory,* 2nd Ed. Boston: Allyn & Bacon.

Hovland, C. I. (1949). Experimental studies in rote-learning theory: VIII. Distributed practice of paired associates with varying rates of presentation. *Journal of Experimental Psychology, 39,* 114-718.

Keller, G. J., Li, Y., Weiss, L. W., & Relyea, G. E. (2006). Contextual interference effect on acquisition and retention of pistol-shooting skills. *Perceptual & Motor Skills, 103,* 241-252.

Kratochwill, T. R., Demuth, D. M., & Conzemius, W. C. (1977). The effects of overlearning on preschool children's retention of sight vocabulary words. *Reading Improvement, 14,* 223-228.

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

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

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59,* 277-287.

Lee, J., Grigg, W. S., & Dion, G. S. (2007). *The nation's report card: Mathematics 2007*. Retrieved September 29, 2007, from http://nces.ed.gov/.

Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of the varied context and varied encoding hypotheses. *Memory & Cognition, 4,* 741-746.

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

Mumford, M. D., Costanza, D. P., Baughman, W. A., Threlfall, K. V., & Fleishman, E. A. (1994). Influence of abilities on performance during practice: Effects of massed and distributed practice. *Journal of Educational Psychology, 86*, 134-144.

National Center for Education Statistics (2007). *National assessment of educational progress: The nation's report card.* Retrieved December, 12, 2007 from http://nces.ed.gov/nationsreportcard/mathematics/

Raajimakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27*, 431-452.

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

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

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal of Research for Mathematics Education, 32,* 368-398.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346-362.

Rittle-Johnson, B. & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99,* 561-574.

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

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

Rohrer, D., Taylor, K., Pashler, H., Cepeda, N. J., & Wixted, J. T. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*, 361-374.

Saxon, J. (1997). *Algebra 1* (3rd Ed.). Norman: Saxon Publishers.

Saxon Elementary School Math. (2007). *What Works Clearinghouse.* U.S. Department of Education.

Saxon Middle School Math. (2007). *What Works Clearinghouse.* U.S. Department of Education.

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

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

Silver, E. A. (2003). Invited Commentary: Lessons learned from examining mathematics teaching around the world. *Education Statistics Quarterly, 5,* 20-23.

Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition, 19,* 159-167.

Toppino, T. C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12,* 118-122.

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

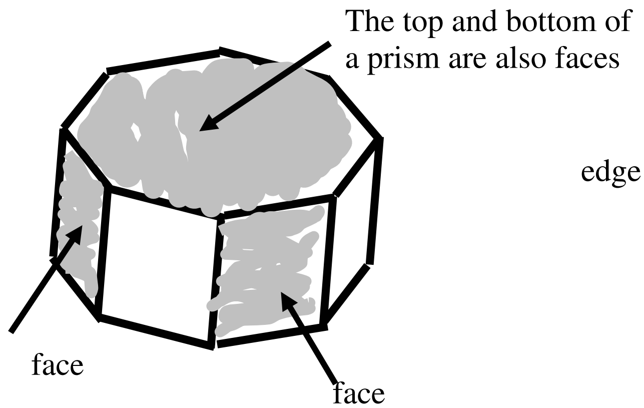
van Merrienboer, J. J. G., & de Croock, M. B. M. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills, 84,* 784-786.

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

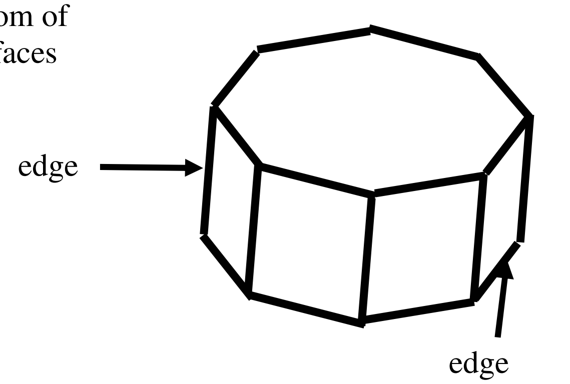
Wu, H. (1999). Basic skills versus conceptual understanding: A bogus dichotomy in mathematics education. *American Educator/American Federation of Teachers,* 1-7

Xin, Y.P. (2007). Word problem solving tasks in textbooks and their relation to student performance. *The Journal of Educational Research (Washington, D.C.), 100,* 347-359.

Appendices

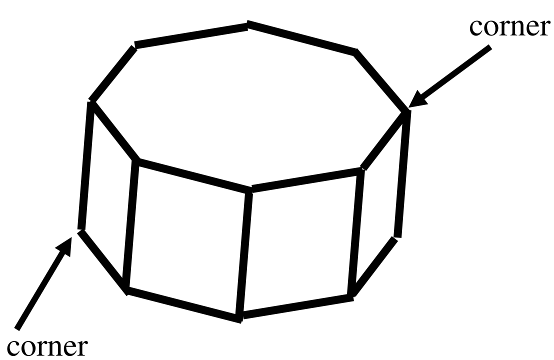


definition: the flat surfaces on the outside of a prism.
formula: faces = b + 2



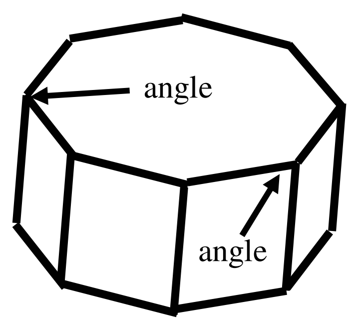
definition: the place where two faces meet.
formula: edges = b x 3

CORNERS



definition: the place where three edges meet.
formula: corners = b x 2

ANGLES



definition: the point where two edges of the same face meet.
formula: angles = b x 6

Appendix B: Example Formula-Learning and Learning-to-Substitute Problems

Example Formula-Learning Problem

On Screen	*On Paper*
1. face	1. <u>b + 2</u> 2. _____ 3. _____ 4. _____

The name of a prism feature appeared on the screen and students wrote the correct formula for that feature on their paper

Example Learning-to-Substitute Problem

Problem	*Solution*
1. A prism with 5 base sides has how many faces? formula: ________ replace: ________	1. A prism with 5 base sides has how many faces? formula: <u>b + 2</u> replace: <u>5 + 2</u>

Students solved each problem by first writing the correct formula for the feature presented and then substituting the number of base sides for *b*.

Appendix C: Example Filler Tasks

Sample Decoding Exercise

Question

What stays hot even after it is put into a refrigerator?

A	E	H	O	P	R	T
5	4	9	7	13	25	14

___   ___ ___ ___
5   9 7 14

___ ___ ___ ___ ___ ___
13 4 13 13 4 25

Solution

A   H O T
5   9 7 14

P E P P E R
13 4 13 13 4 25

Students found the answer to each riddle by using the code provided to fill in the blanks above each number.

Sample Word Search

Question

Circle the words Bee, Cat, and Hat.

C A E C B E A B H
B A H E E E H A T
E H T E E A C E E

Solution

C A E C B E A B H
B A H E E E H A T
E H T E E A C E E

Students circled the words they were asked to find.

About the Author

Kelli Taylor received a Bachelor's Degree in Psychology in 2001 and a M.A. Degree in Psychology in 2004, both from the University of South Florida. During her time in graduate school, she taught an introductory statistics course and coauthored three publications.