

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

The Effects of Overlearning and Distributed Practice on the Retention of Mathematics Knowledge

Doug Rohrer and Kelli Taylor
University of South Florida

In two experiments, 216 college students learned a mathematical procedure and returned for a test either one or four weeks later. In Experiment 1, performance on the four-week test was virtually doubled when students distributed 10 practice problems across two sessions instead of massing the same 10 problems in one session. This finding suggests that the benefits of distributed practice extend to abstract mathematics problems and not just rote memory cognitive tasks. In Experiment 2, students solved 3 or 9 practice problems in a single session, but this manipulation had no effect on either the one-week or four-week test. This result is at odds with the virtually unchallenged support for the strategy of continuing practice beyond the point of mastery in order to boost long-term retention. The results of both experiments suggest that the organization of practice problems in most mathematics textbooks is one that minimizes long-term retention.

Perhaps no mental ability is more important than our capacity to learn, but the benefits of learning are lost once the material is forgotten. Such forgetting is particularly common for knowledge acquired in school, and much of this material is lost within days or weeks of learning. Thus, the identification of learning strategies that extend retention would prove beneficial to students and any others who wish to retain information for meaningfully long periods of time. Toward this aim, the two experiments presented here examined how the retention of a moderately abstract mathematics procedure was affected by variations in either the total amount of practice or the scheduling of this practice.

Specifically, the two learning strategies assessed here are known as overlearning and distributed practice. By an *overlearning* strategy,

a student first masters a skill and then immediately continues to practice the same skill. For example, after a student has studied a set of 20 vocabulary words until each definition has been correctly recalled once, any immediate further study of these definitions entails an overlearning strategy. Overlearning is particularly common in mathematics education because many mathematics assignments include a dozen or more problems of the same type. A *distributed* or *spaced practice* strategy requires that a given amount of practice be divided across multiple sessions and not *massed* into just one session. For example, once a mathematics procedure has been taught to students, the corresponding practice problems can be massed into one assignment or distributed across two or more assignments. As detailed in the general discussion, the practice problems in most mathematics textbooks are arranged so that students rely on overlearning and massed practice.

The strategies of distributed practice and overlearning are not complementary, and the two strategies cannot be compared directly. Instead, distributed practice is the complement of massed practice, and the comparison of these two

Address correspondence to Doug Rohrer, Department of Psychology, PCD4118G, University of South Florida, Tampa, FL 33620, USA. E-mail: drohrer@cas.usf.edu. We thank Christina Coghlan, Tammy Logan, Kristina Martinez, Erica Porch, and Courtney Rosene for their assistance with data collection. This research was supported by two grants from the Institute of Education Sciences, U.S. Department of Education, to Hal Pashler and Doug Rohrer (R305H020061, R305H040108). The opinions expressed are those of the authors and do not necessarily represent the views of the Institute of Education Sciences.

strategies requires that the total amount of practice be held constant. For example, one group of students might divide 10 problems across two sessions while another group solves all 10 problems in the same session (as in Experiment 1). By contrast, assessing the benefits of overlearning requires a manipulation of the total amount of practice given within a single session. Thus, one group might be assigned three problems while another is assigned nine problems (as in Experiment 2). Thus, because overlearning and distributed practice are orthogonal and not complementary, it is logically possible that neither, both, or just one of these strategies could benefit long-term retention. Naturally, both strategies have been the focus of numerous previous studies, but the following review of the research literature reveals caveats, gaps, and inconsistencies with regard to the benefits of each strategy for conceptual mathematics tasks.

Overlearning

An overlearning experiment requires a manipulation of the total amount of practice within a single session, so that one condition includes more practice than another. Numerous experiments have shown that this increase in practice can raise subsequent test performance (e.g., Bromage & Mayer, 1986; Earhard, Fried, & Carlson, 1972; Gilbert, 1957; Kratochwill, Demuth, & Conzemius, 1977; Krueger, 1929; Postman, 1962; Rose, 1992). This benefit of overlearning is also supported by a meta-analysis by Driskell, Willis, and Cooper (1992), who examined 51 comparisons of overlearning versus learning-to-criterion in experiments using cognitive tasks and found a moderately large effect of overlearning on a subsequent test ($d = .75$). By these data, it is not surprising that overlearning is a widely advocated learning strategy (e.g., Fitts, 1965; Foriska, 1993; Hall, 1989; Jahnke & Nowaczyk, 1998).

Yet a closer review of the empirical literature reveals that the benefits of overlearning on subsequent retention may not be long lasting. This is because most previous overlearning experiments have employed a relatively brief *retention interval* (RI), which is the duration between the learning session and the test. For

instance, in the Driskell et al. meta-analysis described above, only 7 of the 51 comparisons relied on a retention interval of more than one week, and the longest was 28 days. Moreover, the largest effect sizes were observed for retention intervals lasting less than one hour, and Driskell et al. astutely observed that benefits of overlearning declined sharply with retention interval.

The possibility that the benefits of overlearning may dissipate with time is also supported by several overlearning experiments including an explicit manipulation of retention interval. For example, in Experiment 1 of Reynolds and Glaser (1964), some students studied biology three times as much as others, and the high studiers recalled 100% more than the low studiers after 2 days but just 7% more after 19 days. A similar decline in the test score benefits of overlearning was observed in two recent experiments reported by Rohrer, Taylor, Pashler, Wixted, and Cepeda (2005).

Finally, it appears that the benefits of overlearning are especially unclear in mathematics learning because, to our knowledge, every previously published overlearning experiment relied solely on verbal memory tasks. Moreover, virtually all of these tasks required only rote memory. Thus, the results of these experiments may not generalize to abstract mathematical tasks. This gap in the literature is surprising in light of the heavy reliance on overlearning by students in mathematics courses, as further detailed in the general discussion. In summary, because of the uncertainty surrounding the long-term benefits of overlearning and the apparent absence of overlearning experiments using mathematics tasks, it is unclear whether overlearning is efficient or even effective when long-term retention is the aim.

Distributed Practice

When practice is *distributed* or *spaced*, a given amount of practice is divided across multiple sessions and not *massed* into one session. The duration of time between learning sessions is the *inter-session interval* (ISI). For example, if 10 math problems are divided across two sessions separated by one week, the ISI equals one week. By contrast, massed practice

entails an ISI of zero. When practice is distributed, the retention interval refers to the duration between the test and the *most recent* learning session. For example, if a concept is studied on Monday and Thursday and tested on Friday, the RI equals one day. (Incidentally, while the benefits of distributing practice *across* sessions is the focus of the present paper, one can also distribute practice *within* a session when multiple presentations are separated by unrelated tasks, e.g., Greene, 1989; Toppino, 1991).

Distributed practice often yields greater test scores than massed practice, and this finding is known as the *spacing effect* (e.g., Baddeley & Longman, 1978; Cull, 2000; Bjork, 1979, 1988; Bloom & Shuell, 1981; Carpenter & DeLosh, 2005; Dempster, 1989; Fishman, Keller, & Atkinson, 1968; Seabrook, Brown, & Soly, 2005). At very brief retention intervals, however, spaced practice may be no better or even worse than massed practice (e.g., Bloom & Shuell, 1981; Glenberg & Lehman, 1980; Krug, Davis, & Glover, 1990). By this result, students are behaving optimally when they mass their learning into one session just prior to an exam *if* they do not need the information after the exam. At longer retention intervals, though, the benefits of spacing are often sizeable, and many researchers have therefore argued that students should rely more heavily on distributed practice in order to extend retention (Bahrck, Bahrck, Bahrck, & Bahrck, 1993; Bjork, 1979; Bloom & Shuell, 1981; Dempster, 1989; Schmidt & Bjork, 1992; Seabrook, Brown, & Soly, 2005).

Even at longer retention intervals, though, the benefits of spacing are more equivocal for tasks that require abstract thinking rather than only rote memory. In a recent meta-analysis of 112 comparisons of distributed and massed practice, Donovan and Radosevich (1999) found that the size of the spacing effect declined sharply as task complexity increased from low (e.g., rotary pursuit) to average (e.g., word list recall) to high (e.g., puzzle). By this finding, the benefits of spaced practice may be muted for mathematical tasks that require more than rote memory. As for mathematical tasks that do rely solely on rote memory, a spacing effect has been demonstrated. For example, Rea and

Modigliani (1985) observed a spacing effect with young children who were asked to memorize five multiplication facts (e.g., $8 \times 5 = 40$). While such facts are certainly useful, the present study focuses on mathematical tasks that require more than rote memory.

A second reason to question the benefits of distributed practice for non-rote mathematics learning is given by the design of three previous mathematics learning experiments that are often cited as evidence of a spacing effect with a non-rote mathematics task. In these experiments, the retention interval was shorter for Spacers than for Massers, and this confound undoubtedly benefited the Spacers. For instance, in Grote (1995), the Massers learned only on Day 1 while the Spacers' learning continued from Day 1 through Day 22. Yet every student was tested on Day 36, which produced a 35-day RI for the Massers and a 14-day RI for the Spacers. This undoubtedly benefited the spacers. The same confound between ISI and RI occurred in the two mathematics learning experiments reported by Gay (1973). We are not aware of any previously published, non-confounded experiment that examines how the distribution of practice affects the retention of a non-rote mathematics task.

There are, however, non-experimental studies that have found benefits of distributed practice of mathematics. Most notably, perhaps, Bahrck and Hall (1991) assessed the retention of algebra and geometry by people who had taken these courses between 1 and 50 years earlier. A regression analysis showed that retention was positively predicted by the number of courses requiring the same material. For example, much of the material learned in an algebra course reappears in an advanced algebra course, and the completion of both courses therefore provides distributed practice of this overlapping material. Thus, the results suggest that distributed practice may be beneficial in mathematics learning. This possibility is assessed here with a controlled experiment, albeit with retention intervals that are measured in weeks rather than years.

Task

In the experiments reported here, college students learned to calculate the number of unique orderings (or permutations) of a letter

sequence with at least one repeated letter. For example, the sequence *abbcc* has 60 permutations, including *abbcb*, *abcbb*, *bbacc*, and so forth. The solution is given by a formula that is presented and illustrated in the Appendix. While the number of permutations can also be obtained by listing each possible permutation, this alternative approach fails when the number of permutations is large and the problem must be completed in a relatively brief amount of time (as in the present study). No student saw a given letter sequence more than once. A total of 17 different sequences were used in the two experiments, and these are listed in the Appendix.

Base Rate Survey

We assessed the base rate knowledge of this task for our experiment participants by giving a brief test to a sample of students drawn from the participant pool used in Experiments 1 and 2. We expected that few if any of the students would be able to perform the task because we have never encountered this particular kind of permutation problem in an undergraduate level textbook.

Method

Participants. The sample included 50 undergraduates at the University of South Florida. These included 43 women and 7 men. None participated in Experiments 1 or 2.

Procedure. After a brief demographic survey, students were given three minutes to solve the following three problems: *aabbbb*, *aaabbb*, and *abcccc*. These sequences yield answers of 21, 35, and 42, respectively.

Results and Discussion

Not surprisingly, none of the students correctly answered any of the three problems. Many of them attempted to solve the problems by listing every permutation, but none exhibited knowledge of the necessary formula. Thus, this mathematics procedure appears to be unknown to the participant pool used in Experiments 1 and 2. Furthermore, even if some participants in these experiments did possess any relevant knowledge before the experiment, the use of random assignment ensured that their presence would not be a confounding variable.

Experiment 1

The first experiment examined the benefit of distributing a given number of practice problems across two sessions rather than massing the same practice problems into one session. As shown in Figure 1A, the Spacers attempted 5 problems in each of two sessions separated by one week, whereas the Massers attempted the same 10 problems in session two. Each group received a tutorial immediately before their first practice problem, and the students were tested one or four weeks after their final practice problem.

Method

Participants. All three sessions were completed by 116 undergraduates at the University of South Florida. The sample included 95 women and 21 men. An additional 39 students completed the first session but failed to show for either the second or third session. None of the students participated in Experiment 2.

Design. There were two between-subjects variables: Strategy (Space or Mass) and Retention Interval (1 or 4 weeks). Thus, each student was randomly assigned to one of four groups: Spacers with 1-week RI, Spacers with 4-week RI, Massers with 1-week RI, and Massers with 4-week RI.

Procedure. The students attended three sessions. At the beginning of the first session, each student was randomly assigned to one of the four conditions listed above. Students were not told what tasks awaited them in future sessions. It is not known whether some students practiced the procedure outside of the experimental sessions, although there was no extrinsic reward for test performance. If any self review did occur, we know of no reason why its prevalence would vary among Spacers and Massers.

All students completed a tutorial, two practice sets, and a test. Students read the tutorial immediately before their first practice set. The tutorial included two pages of instruction (3 min) and written solutions to Problem 11 (3 min) and Problem 12 (2 min) of the Appendix. The first practice set included Problems 1 – 5 of the Appendix, in that order, and the second practice

set included Problems 6 – 10 of the Appendix, in that order. Each problem appeared in a booklet with only one problem per page. Students were allotted 45 s to solve each problem. Immediately after each attempt, students were shown the solution for 15 s before immediately beginning the next problem.

The first two sessions were separated by one week. The Spacers completed the first five-problem practice set in session one and the second five-problem practice set in session two. The Massers completed both practice sets in session two, without any delay between the two sets.

One or four weeks after the second session, every student returned for a test. The test included the five test problems listed in the Appendix in the order shown. The five test problems were presented simultaneously, and students were allotted five minutes to solve all five problems. No feedback was given during the test.

Results

Learning. The tutorial was sufficiently effective, as evidenced by performance on the five problems given to every student immediately after the tutorial. Of the 116 students, 65 scored a perfect five, 27 scored four, 12 scored three, 6 scored two, 3 scored one, and 3 scored zero. All further analyses excluded the students who correctly answered zero ($n = 3$) or one ($n = 3$) of these five problems. It appears that the three students with scores of one were guessing, as four of the five correct answers were multiples of ten. Notably, this inclusion criterion of at least two correct is based solely on performance during the first session because the additional reliance on performance during the second practice set (which was delayed for the Spacers) would have confounded the experiment. Naturally, the Massers and Spacers performed equivalently on the first practice set because the procedures for these groups did not diverge until *after* the first practice set was complete. Specifically, the Massers averaged 88% (SE = 2.3%) and the Spacers averaged 87% (SE = 2.5%), $F < 1$.

However, for the second set of five practice problems, the Massers averaged 94%

(SE = 1.6%) while the Spacers averaged only 85% (SE = 3.2%), $F(1, 108) = 6.79, p < .05, \eta_p^2 = .06$. This difference was due to very poor performance by a subset of Spacers who apparently forgot the procedure during the one week inter-session interval. Thus, the one-week ISI introduced a disadvantage for the Spacers, but this worked *against* the spacing effect rather than for it.

Test. The mean percentage accuracy for the five test problems is shown in Figure 1B. As illustrated, the Spacers and Massers were not reliably different at the 1-week RI, but the Spacers sharply outscored the Massers at the 4-week RI. This parity at one week caused the main effect of Strategy (Space vs. Mass) to fall short of statistical significance, $F(1, 106) = 3.67, p = .06, \eta_p^2 = .03$. Not surprisingly, the main effect of RI was reliable, $F(1, 106) = 12.92, p < .001, \eta_p^2 = .11$. The reliance of the spacing effect on retention interval was evidenced by an interaction between ISI and RI, $F(1, 106) = 7.21, p < .01, \eta_p^2 = .06$. This pattern was further confirmed by Tukey tests showing that the difference between Spacers and Massers was reliable at the 4-week RI ($p < .05$) but not the 1-week RI. These post hoc tests also showed that the difference between the one- and four-week test scores was significant for the Massers ($p < .05$) but not the Spacers.

Discussion

For the longer retention interval of four weeks, the distribution of 10 practice problems across two sessions was far more useful than the massing of all 10 problems in the same session. Thus, these data provide an instance of the spacing effect for a non-rote mathematics learning task in a non-confounded experiment. At the one-week retention interval, there was no reliable difference between the two strategies among students who were tested only one week after learning. This result is consistent with previous findings demonstrating no spacing effect or even massing superiority at sufficiently short retention intervals, as described in the introduction. Despite this ambiguity after one week, though, the spacing superiority after four weeks suggests that long-term retention, which is

the focus of the present paper, is better achieved by distributing practice problems across sessions.

Experiment 2

The second experiment assessed the effect of overlearning on retention by varying the number of practice problems within a single session. The Hi Massers attempted 9 practice problems, whereas the Lo Massers attempted only 3 practice problems (as detailed in Figure 1C). Thus, the Hi Massers relied heavily on overlearning. Because this increase in the number of practice problems produced a concomitant increase in time devoted to practice, this manipulation is effectively a manipulation of total practice time. Students were tested either one or four weeks later. As detailed in the introduction, many researchers have found benefits of overlearning on a subsequent test, but the bulk of these experiments relied on relatively brief retention intervals.

Method

Participants. All three sessions were completed by 100 undergraduates at the University of South Florida. The sample included 83 women and 17 men, and none participated in Experiment 1. An additional 17 students completed the first session but failed to show for the second session.

Design. We manipulated two between-subjects variables: Practice Amount (Hi or Lo) and Retention Interval (1 or 4 weeks). Thus, each student was randomly assigned to one of four groups: Hi Massers with 1-week RI, Hi Massers with 4-week RI, Lo Massers with 1-week RI, and Lo Massers with 4-week RI.

Procedure. Each student attended two sessions, separated by one or four weeks. At the beginning of the first session, each student was randomly assigned to one of the four conditions listed above. Each student then observed a tutorial consisting of screen projections that included the complete solutions to Problems 10, 11, and 12 of the Appendix, in that order. Immediately after this tutorial, all students began the practice problems. Each Hi Masser was given nine problems (which were Problems 1 through 9 of the Appendix), and each Lo Masser was assigned three of these nine problems. The three problems assigned to each Lo Masser varied, so

that each of the nine problems was presented equally often. This was done to equate the selection of practice problems given to Lo and Hi Massers. In addition, the nine problems given to the Hi Massers were presented in one of three different orders so that the *first* three problems corresponded to the *only* three problems given to the same number of Lo Massers. The other aspects of the procedure, including the five-problem test, were the same as those in Experiment 1.

Results

Learning. The tutorial was again sufficient to produce learning, as demonstrated by students' performance on the three practice problems completed immediately after the tutorial. Specifically, 65 of the 100 students correctly answered all three problems, 23 scored two, 10 scored one, and 2 scored zero. As in Experiment 1, students with scores of zero or one were excluded from further analysis.

As expected, there was no reliable difference between Hi and Lo Massers on the first three problems because these two groups underwent the same procedure until *after* these three problems were completed. Specifically, the Hi Massers averaged 90% (SE = 2.3%) and the Lo Massers averaged 88% (SE = 2.3%), $F < 1$. For the additional six practice problems given only to the Hi Massers, accuracy averaged 95% (SE = 1.3%).

Test. As shown in Figure 1D, there was virtually no difference between the Hi and Lo Massers on either the one- or four-week test. Consequently, an analysis of variance revealed no main effect of Practice Amount ($F < 1$) and no interaction between Practice Amount and Retention Interval ($F < 1$). Not surprisingly, the main effect of retention interval was significant, $F(1, 84) = 33.16, p < .001, \eta_p^2 = .28$.

Discussion

The increase in the number of practice problems given during a single learning session had virtually no effect on subsequent test scores at either retention interval. This null effect of overlearning is not well explained by a lapse in attention by the Hi Massers during their additional six practice problems because these

problems were solved with 95% accuracy. Thus, as fully described in the general discussion, these results provide no support for the oft cited claim that overlearning boosts long-term retention and therefore cast doubt on the utility of mathematics assignments that include many problems of the same type.

General Discussion

The two experiments assessed the benefits of distributed practice and overlearning on subsequent test performance. In Experiment 1, distributing 10 practice problems across two sessions instead of massing all 10 problems in the same session had no effect on one-week test scores but virtually doubled four-week test scores. In Experiment 2, increasing the number of problems solved in a single session from three to nine had virtually no effect on test scores on either the one- or four-week test. In brief, the extra effort devoted to additional problems produced no observable benefit, whereas the distribution of a given number of number of practice problems produced benefits without any extra effort.

The results of previous experiments have provided little support for distributed practice with non-rote mathematics tasks. As detailed in the introduction, three previously published experimental findings that are cited as instances of a spacing effect with a non-rote mathematics task are, in fact, confounded in favour of the spacing effect. The results of Experiment 1, however, suggest that the superiority of distributed practice over massed practice extends to these more abstract cognitive tasks. Consequently, we concur with those authors who have urged greater reliance on distributed practice as a means of boosting long-term retention (Bahrick & Hall, 1991; Baddeley & Longman, 1978; Bjork, 1979, 1988; Dempster, 1989; Reynolds & Glaser, 1964; Schmidt & Bjork, 1992; Seabrook, Brown, & Soly, 2005).

With regard to overlearning, however, the present results strongly conflict with the numerous claims about its utility as a learning strategy (e.g., Fitts, 1965; Foriska, 1993; Hall, 1989; Jahnke & Nowaczyk, 1998). Indeed, the strategy of overlearning is widely advocated. As Jahnke and Nowaczyk advised, "Practice should

proceed well beyond that minimally necessary for an immediate, correct first reproduction" (p. 181). Fitts concluded that, "The importance of continuing practice beyond the point in time where some (often arbitrary) criterion is reached cannot be overemphasized" (p. 195). And Hall wrote, "The overlearning effect would appear to have considerable practical value since continued practice on material already learned to a point of mastery can take place with a minimum of effort, and yet will prevent significant losses in retention" (p. 328). In contrast to these conclusions, the results of Experiment 2 revealed no effect of overlearning on retention.

Conceptually, the minimal effect of overlearning on retention can be interpreted as an instance of diminishing returns. That is, with each additional amount of practice devoted to a single concept, there is an ever smaller increase in test performance. Thus, after the initial exposure to a concept, the first one or two practice problems might yield a large increase in a subsequent test score. Yet each additional practice problem provides an ever smaller gain until, ultimately, any additional practice within the same session will yield very little gain.

It should be noted, though, that a small amount of overlearning may be useful if overlearning is strictly defined as any practice beyond one correct problem. By this definition, overlearning occurs even when a student correctly solves only two or three problems, which means that even the *Lo Massers* in Experiment 2 relied on a small amount of overlearning. Consequently, the results of Experiment 2 do not support the extreme view that students should be assigned only one problem of each type in a given session. Instead, the results suggest that students who correctly solve several problems of the same kind have little to gain by working more problems of the same type within the same session. After these first problems are solved correctly, students could devote the remainder of the practice session to problems drawn from previous lessons in order to reap the benefits of distributed practice.

A final caveat concerns two important limitations on the extent to which the present

finding will generalize. First, all of the participants in the present experiments were college students, and it is not known whether the results of both experiments would have been observed with much younger students. However, we suspect that young children would provide qualitatively similar results because such parallels have been observed in previous distributed practice experiments (e.g., Seabrook, Brown, & Solity, 2005; Toppino, 1991). Second, because the experiments reported here relied on a test that required students to solve problems identical to those presented during the practice session (albeit with different numerical values), it is unknown whether the benefit of distributed practice or the futility of overlearning would have occurred if the test had required students to apply their previous learning in novel ways (i.e., assessed what is known as transfer).

The Organization of Practice Problems in Mathematics Textbooks

Many mathematics textbooks rely on a format that fosters both overlearning and massed practice. In these textbooks, virtually all of the problems for a given topic appear in the assignment that immediately follows the lesson on that topic. This format fosters overlearning because each assignment includes many problems of the same kind. This format also fosters massed practice because further problems of the same kind are rarely included in subsequent assignments. As an illustration, we examined every problem in the most recent editions of four textbooks in pre-algebra mathematics or introductory algebra that are very popular in the United States. The proportion of the problems within each assignment that corresponded to the immediately preceding lesson, when averaged across assignments, equalled between 75% and 92% for the four books. Thus, the format of these practice sets facilitates overlearning and massing.

Fortunately, there is an alternative format that minimizes overlearning and massed practice while emphasizing distributed practice, and it does not require an increase in the number of assignments or the number of problems per assignment. With this distributed-practice format, each lesson is followed by the usual

number of practice problems, but only a few of these problems relate to the immediately preceding lesson. Additional problems of the same type then appear perhaps once or twice in each of the next dozen or so assignments and once again after every fifth or tenth assignment thereafter. In brief, the number of practice problems relating to a given topic is no greater than that of typical mathematics textbooks, but the temporal distribution of these problems is increased dramatically.

While the distributed practice format should improve retention, it might also prove more challenging than the massed practice format. With a massed practice format, students who have solved the first few problems have little difficulty with the remaining problems of the same type. Indeed, they merely need to repeat the procedure. A distributed practice format, however, ensures that each practice set includes many different challenges. Thus, the challenge and long-term returns of a distributed practice format provide an example of what Bjork and his associates have called a “desirable difficulty” (Christina & Bjork, 1991; Schmidt & Bjork, 1992). Despite this difficulty, however, some students might find the mixture of problems more interesting than a group of similar problems.

A distributed practice format is used in the Saxon series of mathematics textbooks (e.g., Saxon, 1997). While numerous non-controlled studies have compared Saxon textbooks to other textbooks, we are not aware of any published, controlled experiments with these textbooks. However, there may be little information to be gained from an experiment in which students are randomly assigned to a condition that uses a Saxon or non-Saxon textbook because the numerous differences between two such textbooks would confound the experiment. For example, if such an experiment did show superior retention for users of the Saxon textbook, it is logically possible that this benefit was the result of textbook features other than the distributed practice format. (Neither author has had any affiliation with Saxon Publishers; however, the first author is a former mathematics

teacher who has used both Saxon and non-Saxon mathematics textbooks in the classroom.)

Perhaps a more informative experiment would compare two groups of students who underwent instruction programs that differed only in the temporal distribution of practice problems. For example, a class of students could be divided randomly into two groups, with each group participating in the same class activities. Every student would also receive a packet that included the same lessons in the same order. The selection of practice problems would also be identical, but the practice problems of each type would be distributed or massed.

Textbook publishers could adopt a distributed-practice format with little trouble or cost. They would merely rearrange the practice problems in the next edition of their textbooks, regardless of whether the lessons are changed as well. Oddly, practice problems typically receive relatively little attention from publishers and textbook authors, and the practice problems are often written by sub-contracted writers. Yet the practice sets are at least as important as the lessons. In fact, as many mathematics teachers will attest, a majority of their students never read the lessons and instead devote all of their individual effort to the practice sets.

In addition to the implications for textbook design, the benefits of distributed practice are equally applicable to the design of the algorithms used in computer-aided instruction (CAI). Unlike textbooks, the programs can provide individualized training and error-contingent feedback, and an increasing number of educators and agencies have urged greater reliance on such technologies (e.g., Department for Education and Skills, United Kingdom, 2003). Yet virtually all currently available CAI programs are designed to foster learning rather than retention. Of course, such programs could be easily adapted to incorporate distributed rather than massed practice, and students' compliance to a distributed practice schedule could be verified by ensuring that the

program record the date of each problem attempt.

In addition to its effect on retention, a distributed-practice format can also facilitate learning because it allows students ample time to master a particular skill. For instance, if a student is unable to solve the problems of a given type in a single lesson, a distributed-practice format will provide further opportunities throughout the year.

Conclusion

The results of Experiment 1 suggest that the retention of mathematics is markedly improved when a given number of practice problems relating to a topic are distributed across multiple assignments and not massed into one assignment. Moreover, this benefit of distributed practice can be realized without increasing the number of practice problems included in a practice set typical of most mathematics textbooks. Specifically, rather than require students to work far more than just a few problems of the same kind in the same session, which had no effect in Experiment 2, each practice set could instead include problems relating to the most recent topic as well as problems relating to previous topics. This distributed practice format could be easily adopted by the authors of textbooks and CAI software.

Any resulting boost in students' mathematics retention might greatly improve the mathematics achievement, and there is little doubt that the mathematics skills of most students need improving. In one recent report on mathematics achievement, less than one third of a sample of U.S. students received a rating of "at or above proficient" (Wirt et al., 2004). Such reports often lead people to conclude that students are not learning, but it may be that many mathematical skills and concepts are learned but later forgotten. The prevalence of such forgetting may partly reflect the widespread reliance on practice schedules that proved to be the worst strategies in the experiments reported here.

References

- Baddeley, A. D., & Longman, D. J.A. (1978). The influence of length and frequency of training sessions on the rate of learning to type. *Ergonomics*, *21*, 627-635.
- Bahrack, H. P., Bahrack, L.E., Bahrack, A.S., & Bahrack, P. E. (1993). Maintenance of foreign-language vocabulary and the spacing effect. *Psychological Science*, *4*, 316-321.
- Bahrack, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, *120*, 20-33.
- Bjork, R. A. (1979). Information-processing analysis of college teaching. *Educational Psychologist*, *14*, 15-23.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M.M. Gruneberg, P.E., Morris, & R.N. Sykes (Eds.). *Practical aspects of memory II* (pp. 391-401). London: Wiley.
- 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.
- 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.
- Carpenter, S.K., & DeLosh, E.L. (2005). Application of the testing and the spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619-636. DOI: 10.1002/acp.1101
- 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.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215-235.
- Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*, 309-330.
- Department for Education and Skills, United Kingdom (2003). *Fulfilling the potential*. (DfES publication 0265 / 2003). DfES Publications.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *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.
- Earhard, B., Fried, C., & Carlson, G. (1972). Interference, overlearning, and anticipation time. *Journal of Experimental Psychology*, *94*, 345-347.
- Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling. *Journal of Educational Psychology*, *59*, 290-296.
- Fitts, P. M. (1965). Factors in complex skill training. In R. Glaser (Ed.), *Training research and education* (pp. 177-197). New York: Wiley.
- Foriska, T. J. (1993). What every educator should know about learning. *Schools in the Middle*, *3*, 39-44.
- 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.
- Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition*, *8*, 528-538.
- 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.

- Grote, M. G. (1995). Distributed versus massed practice in high school physics. *School Science and Mathematics, 95*, 97-101.
- Hall, J. F. (1989). *Learning and memory, 2nd Ed.* Boston: Allyn & Bacon.
- Jahnke, J.C., & Nowaczyk, R. H. (1998). *Cognition*. Upper Saddle River, NJ: Prentice Hall.
- Kratochwill, T., 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.
- Krug, D., Davis, T. B., & Glover, J. A. (1990). Massed versus distributed repeated reading – a case of forgetting helping recall. *Journal of Educational Psychology, 82*, 366-371.
- Postman, L. (1962). Retention as a function of degree of overlearning. *Science, 135*, 666-667.
- 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.
- Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*, 361-374. DOI: 10.1002/acp.1083
- Rose, R. J. (1992). Degree of learning, interpolated tests, and rate of forgetting. *Memory & Cognition, 20*, 621-632.
- Saxon, J. (1997). *Algebra I (3rd Ed.)*. Norman, OK: Saxon Publishers.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.
- Seabrook, R., Brown, G.D.A., & Solity, J.E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107-122. DOI: 10.1002/acp.1066
- Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition, 19*, 159-167.
- Wirt, J., Choy, S., Rooney, P., Provasnik, S., Sen, A., & Tobin, R. (2004). The Condition of Education 2004 (NCES 2004-077). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Appendix

Students were taught to calculate the number of unique orderings (or permutations) of a letter sequence with at least one repeated letter (e.g., *abbbcc*). For n items and k unique items, the number of permutations equals $n! / (n_1! n_2! \dots n_k!)$, where n_i = number of repetitions of item i . For example, *abbbcc* includes six letters ($n = 6$) and three unique letters ($k = 3$), and the letters a , b , and c appear 1, 3, and 2 times, respectively ($n_1 = 1, n_2 = 3, n_3 = 2$). Thus, by the formula, the number of permutations equals

$$\begin{aligned} & 6! / (1! 3! 2!) \\ & = (6 \times 5 \times 4 \times 3 \times 2 \times 1) / [(1) \times (3 \times 2 \times 1) \times (2 \times 1)] \\ & = (6 \times 5 \times 4) / (2) \\ & = 60. \end{aligned}$$

Tutorial and Learning Session (see the procedures of each experiment for details)

1. abccc	20
2. abcccc	30
3. aabbbbb	21
4. abbcc	30
5. aaabbb	20
6. aabbbb	15
7. aabb	6
8. abbccc	60
9. abccccc	42
10. aabbbbbbb	28
11. abbcccc	105
12. abccccc	56

Test

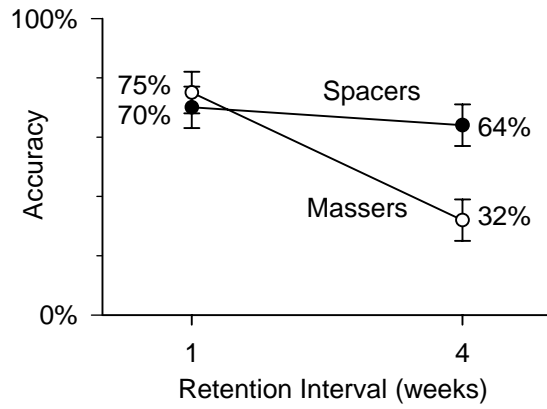
abcc	12
aaabbb	10
aabbcc	90
aaabbbb	35
aaaabbbb	70

Experiment 1

A Learning Procedure

	1 week apart	
	session one	session two
Spacers	5 problems	5 problems
Massers	–	10 problems

B Test Results



Experiment 2

C Learning Procedure

	session one
Hi Massers	9 problems
Lo Massers	3 problems

D Test Results

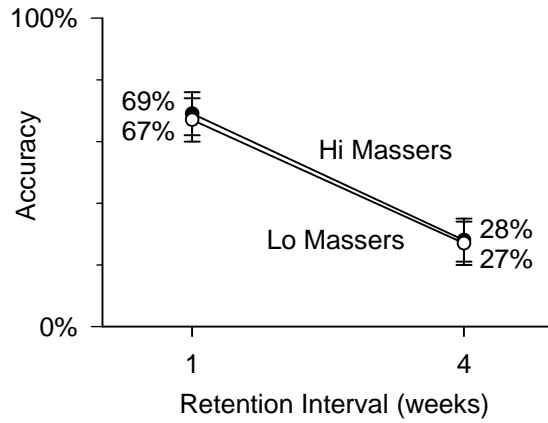


Figure 1. Learning Procedure and Test Results for Experiments 1 and 2. Accuracy represents the mean percentage correct. Error bars reflect plus or minus one standard error.