APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **20**: 1209–1224 (2006)
Published online 6 July 2006 in Wiley InterScience
(www.interscience.wiley.com). DOI: 10.1002/acp.1266

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

DOUG ROHRER* and KELLI TAYLOR

University of South Florida, USA

SUMMARY

In two experiments, 216 college students learned to solve one kind of mathematics problem before completing one of various practise schedules. In Experiment 1, students either massed 10 problems in a single session or distributed these 10 problems across two sessions separated by 1 week. The benefit of distributed practise was nil among students who were tested 1 week later but extremely large among students tested 4 weeks later. In Experiment 2, students completed three or nine practise problems in one session. The additional six problems constituted a strategy known as overlearning, but this extra effort had no effect on test scores 1 or 4 weeks later. Thus, long-term retention was boosted by distributed practise and unaffected by overlearning. Unfortunately, most mathematics textbooks rely on a format that emphasises overlearning and minimises distributed practise. An easily adopted alternative format is advocated. Copyright © 2006 John Wiley & Sons, Ltd.

Perhaps no mental ability is more important than our capacity to learn, but the benefits of learning are mostly lost if 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, any learning strategy must be judged at least in part by students' performance after a non-trivial retention interval (RI), a criterion that is typically ignored in education research. Here, we present two experiments that examined how the retention of a mathematics procedure was affected by variations in the temporal distribution of practise or the total amount of practise.

Specifically, we assessed the learning strategies of distributed practise and overlearning. When practise is *distributed* or *spaced*, a given amount of practise is divided across multiple sessions. For example, once students have learned to solve a mathematics procedure, the corresponding practise problems can be *massed* into one assignment or distributed across multiple assignments. Most mathematics textbooks emphasise massed practise, as most of the problems relating to a given topic typically appear in the same practise set. By an *overlearning* strategy, a student first masters a skill and then immediately continues to practise the same skill. Overlearning is particularly common in mathematics because assignments typically require students to solve many problems of the

*Correspondence to: D. Rohrer, Department of Psychology, PCD 4118G, University of South Florida, Tampa, FL 33620, USA. E-mail: drohrer@cas.usf.edu

Contract/grant sponsor: Institute of Education Sciences, U.S. Department of Education; contract/grant numbers: R305H020061, R305H040108.

same type. Notably, the term overlearning refers to a strategy and not the ultimate degree of mastery. For example, one can master the names of the 12 calendar months without ever using the strategy of overlearning. It is the *strategy* of overlearning that is assessed here and not the utility of mastery.

The strategies of distributed practise and overlearning are not complementary and cannot be compared directly. Instead, distributed practise must be compared to massed practise by holding constant the total amount of practise. In Experiment 1, for example, students worked 10 problems that were either distributed across two sessions or massed into one session. By contrast, the benefits of overlearning are assessed by varying the amount of practise within a given session. In Experiment 2, for example, students worked three or nine problems in the same session. Because distributed practise and overlearning are orthogonal, it is logically possible that neither, both, or just one of these two strategies is beneficial. Naturally, both strategies have been the focus of numerous previous studies, but a review of the literature reveals a few caveats, gaps and confounds.

DISTRIBUTED PRACTISE

In most distributed practise experiments, practise is either massed into a single session or distributed across two sessions separated by a period of time known as the *inter-session interval*. For example, if 10 math problems are divided across two sessions separated by 1 week, the inter-session interval equals 1 week. Importantly, the RI equals the duration between the test and the *most recent* learning session. For example, if a concept is practised on Monday and Thursday and tested on Friday, the RI equals 1 day. (Incidentally, practise can be spaced *within* a session by presenting material on two different trials with at least one intervening unrelated trial, e.g. Carpenter & DeLosh, 2005; Greene, 1989; Toppino, 1991, but the present paper focuses on the effect of distributing practise *across* sessions.)

Subsequent retention is often greater if practise is distributed rather than massed—a finding known as the *spacing effect* (e.g. Baddeley & Longman, 1978; Bloom & Shuell, 1981; Cull, 2000; Fishman, Keller, & Atkinson, 1968; Seabrook, Brown, & Solity, 2005). At brief RIs, however, spaced practise is typically no better or even worse than massed practise (e.g. Bloom & Shuell, 1981; Glenberg & Lehmann, 1980; Krug, Davis, & Glover, 1990). By this latter result, massing just prior to an exam is optimal if the information needs not be retained after the exam. At longer RIs, though, which are of greater interest to educators, the benefits of spacing are often sizeable. Consequently, many authors have urged a greater reliance on distributed practise (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bjork, 1979, 1988; Bloom & Shuell, 1981; Dempster, 1989; Reynolds & Glaser, 1964; Schmidt & Bjork, 1992).

Even at longer RIs, though, the benefits of spacing are questionable for cognitive tasks that are more conceptually demanding than those that require only verbatim recall. In one meta-analysis by Donovan and Radosevich (1999), for instance, the size of the spacing effect declined sharply as conceptual difficulty of the task 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 practise may be muted for many mathematics tasks. Of course, not all mathematics knowledge is conceptual. For example, Rea and Modigliani (1985) found a spacing effect with young children who studied multiplication facts (e.g. $8 \times 5 = 40$). While such concrete facts are certainly useful, the present study examines the benefits of distributed practise for tasks that require more than the verbatim recall of atomised facts.

A second reason to question the benefit of distributed practise for conceptual mathematics tasks is given by a confound in the design of three previous mathematics learning experiments that are sometimes cited as instances of a spacing effect. In each of these experiments, the RI was shorter for Spacers than for Massers, and this confound undoubtedly benefited the Spacers. In Grote (1995), for instance, the Massers practised only on Day 1, while the Spacers' practise continued from Day 1 through Day 22. Yet every student was tested on Day 36, leading to a far shorter RI for Spacers that undoubtedly worked in their favour. The same confound arose in 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 practise affects the retention of a mathematics task that requires more than verbatim recall.

There are, however, non-experimental studies that have examined the effect of distributed practise on mathematics retention. Most notably, perhaps, Bahrick 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 practise of this overlapping material. That such distributed practise is beneficial is assessed here with a controlled experiment, albeit with RIs that are measured in weeks rather than years.

OVERLEARNING

An overlearning experiment requires a manipulation of the total amount of practise within a single session. In a typical overlearning experiment, subjects either quit studying once they achieve a criterion of one correct instance (i.e. learn to criterion) or reach criterion and then immediately continue to study (i.e. overlearn). In comparison to learning-to-criterion, overlearning typically boosts 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 Copper (1992) who examined 51 comparisons of overlearning and learning-to-criterion with cognitive tasks and found a moderately large effect of overlearning on subsequent test scores ($d = 0.75$). Hence, it is not surprising that overlearning is widely advocated (e.g. Fitts, 1965; Foriska, 1993; Hall, 1989; Jahnke & Nowaczyk, 1998).

Yet a closer review of the empirical literature reveals that the long-term benefits of overlearning remain unclear because most overlearning experiments have used relatively brief RIs. For instance, in the Driskell et al. (1992) meta-analysis, only 7 of the 51 comparisons relied on a RI of more than 1 week, and the largest effect sizes were observed for RIs lasting less than 1 hour. In fact, Driskell et al. concluded that RI moderated the benefits of overlearning.

The possibility that the benefits of overlearning may dissipate with time is also supported by three overlearning experiments with an explicit manipulation of RI. In Experiment 1 of Reynolds and Glaser (1964), some students studied biology three times as much as other students, and this threefold increase in study time led to a 100% boost in test scores 2 days later that decreased to just 7% difference 19 days after learning. 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 for mathematics because, to our knowledge, no previously published overlearning experiment has used a mathematics task. Of those that used cognitive tasks (rather than motor tasks), all employed a verbal memory task that almost always required the verbatim recall of either paired associates or a list of words. In brief, it appears that little is known about the effect of overlearning on mathematics retention.

TASK

In the experiments reported here, college students calculated the number of unique orderings (or permutations) of a letter sequence with at least one repeated letter. For example, the sequence *abbbcc* has 60 permutations, including *abbcbc*, *abcbcb*, *bbacbc* and so forth. The solution is given by a straightforward procedure that is illustrated in the Appendix.

This task is moderately abstract, unlike the verbatim recall of an atomised fact (e.g. $8 \times 5 = 40$). In addition, students saw no problem more than once, ensuring that they could not merely memorise the answer for a given letter sequence. Still, like most mathematics procedures, the task is a memory task because students must remember a series of steps.

BASE RATE SURVEY

We assessed the base rate knowledge of this permutation task by testing a sample of students drawn from the participant pool used in Experiments 1 and 2. We expected that none of the students would be able to perform the task because this particular kind of permutation problem appears in virtually no undergraduate level textbooks.

Method

Participants

The students were 50 undergraduates at the University of South Florida. These included 43 women and 7 men, and none participated in Experiments 1 or 2.

Procedure

Each student was given 3 minutes to find the number of permutations for the sequences, *aabbbbb*, *aaabbbb* and *abccccc*. The answers are 21, 35 and 42 respectively.

Results and discussion

None of the students correctly answered any of the three problems, and none of their written solutions exhibited any knowledge of the appropriate procedure. Some students attempted to simply list every permutation but none succeeded. Thus, this mathematics procedure appears to be mostly or entirely unknown to the participant pool used in Experiments 1 and 2. Furthermore, if any of the participants in Experiments 1 or 2 did have any relevant knowledge before the experiment, their presence did not confound the experiment because participants were randomly assigned to condition.

EXPERIMENT 1

The first experiment examined the benefit of distributing a given number of practise problems across two sessions instead of massing the same practise problems into one session. As shown in Figure 1A, the Spacers attempted five problems in each of two sessions separated by 1 week, whereas the Massers attempted the same 10 problems in Session 2. Each group received a tutorial immediately before their first practise problem, and the students were tested 1 or 4 weeks after their final practise problem.

Method

Participants
All three sessions were completed by 116 undergraduates at the University of South Florida. This 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.

Design
Each student was randomly assigned to one of four groups: Spacers with 1-week RI ($n = 29$), Spacers with 4-week RI ($n = 29$), Massers with 1-week RI ($n = 31$) and Massers with 4-week RI ($n = 27$).

Procedure
The students attended three sessions. At the beginning of the first session, each student was randomly assigned to one of the four conditions by the random distribution of paper packets. Students were not told what tasks awaited them in future sessions. It is not known whether some students practised 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 between Spacers and Massers.

The first two sessions were separated by 1 week. The Spacers completed five problems in Session 1 and an additional five problems in Session 2, whereas the Massers completed the same 10 problems in the second session. Thus, once students were assigned to conditions at the beginning of the first session, the Massers departed before learning anything about the permutation task. This meant that the Massers were required to attend the same three sessions as the Spacers, thereby preventing a confound due to differential rates of attrition. That is, if Massers had been required to attend only two sessions, the data for the Massers (but not the Spacers) might have included students who attended the second but not the third session. To the extent that these third-session non-attendees are less capable than their cohorts, this confound would have worked in favour of the spacing effect.

The first practise problem was immediately preceded by a tutorial including two pages of general instructions unrelated to the permutation task (3 minutes) and written solutions to two sample problems, which are listed in the Appendix. The tutorial did not include the general formula because it includes factorial expressions (e.g. $n!$) that we believed would be familiar to some but not all of our participants. Instead, each tutorial example included a two-step solution that excluded factorials and variables, and this procedure is illustrated in the Appendix.

Immediately after the tutorial, students began the 10 practise problems. The order of the 10 problems did not vary, and the specific problems are listed in the Appendix. Each problem appeared on a page by itself within a booklet. Students were allotted 45 seconds to solve each practise problem, and, immediately after each attempt, they were shown the solution for 15 seconds. The solution included the same two procedural steps that were provided in the tutorial examples, as illustrated in the Appendix. The Massers received all 10 practise problems in the second session, whereas the Spacers received the first five problems in Session 1 and the second five problems in Session 2. Notably, the Spacers did not receive any review or tutorial during the second session.

One or four weeks after the second session (in which every student received their last practise problem), students returned for the third and final session to be tested. The test included the five test problems listed in the Appendix in the order shown there. The problems were presented simultaneously, and students were allotted 5 minutes to solve all five problems. No feedback was given during the test.

Results and discussion

Practise problems

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 included only those students who correctly answered two or more of these five problems, thereby eliminating 12 of the 116 students.

Naturally, the Massers and Spacers performed equivalently on the first five practise problems because these two groups underwent identical procedures until *after* they attempted these problems. Specifically, the Massers averaged 88% (SE = 2.3%) and the Spacers averaged 87% (SE = 2.5%), $F < 1$. For the second set of five practise problems, the Massers' average of 94% (SE = 1.6%) significantly exceeded the Spacers' average of 85% (SE = 3.2%), $F\,(1, 108) = 6.79$, $p < 0.05$, $\eta_p^2 = 0.06$. This difference was presumably due to forgetting by the Spacers during the 1-week interval between their two practise sessions.

Test

Mean test accuracy is plotted in Figure 1B. The Spacers and Massers scored about equally on the 1-week test, but a large spacing effect was observed on the 4-week test. An analysis of variance revealed no main effect of spacing because of the 1-week parity, $F\,(1, 106) = 3.67$, $p = 0.06$, $\eta_p^2 = 0.03$, but the main effect of RI was significant, $F\,(1, 106) = 12.92$, $p < 0.001$, $\eta_p^2 = 0.11$. More importantly, the interaction between the size of the spacing effect and the RI was also significant, $F\,(1, 106) = 7.21$, $p < 0.01$, $\eta_p^2 = 0.06$. This effect of RI on the spacing effect was further confirmed by Tukey tests showing that the difference between Spacers and Massers was reliable at the 4-week RI ($p < 0.05$) but not the 1-week RI. The Tukey tests also showed that the decline between the 1-week and 4-week test scores was significant for Massers ($p < 0.05$) but not Spacers.

In summary, the large spacing effect on the 4-week test suggests that the benefits of distributed practise extend to conceptual tasks and are not limited to tasks requiring the verbatim recall of atomised facts. While there was no spacing effect on the 1-week test, this finding is consistent with previously published comparisons of spacing and massing at brief RIs, as described in the introduction. This parity at 1 week is not problematic from a practical viewpoint, though, because the possibility that the spacing effect *grows* with RI would merely be another reason why long-term mathematics retention is better achieved when practise is distributed rather than massed.

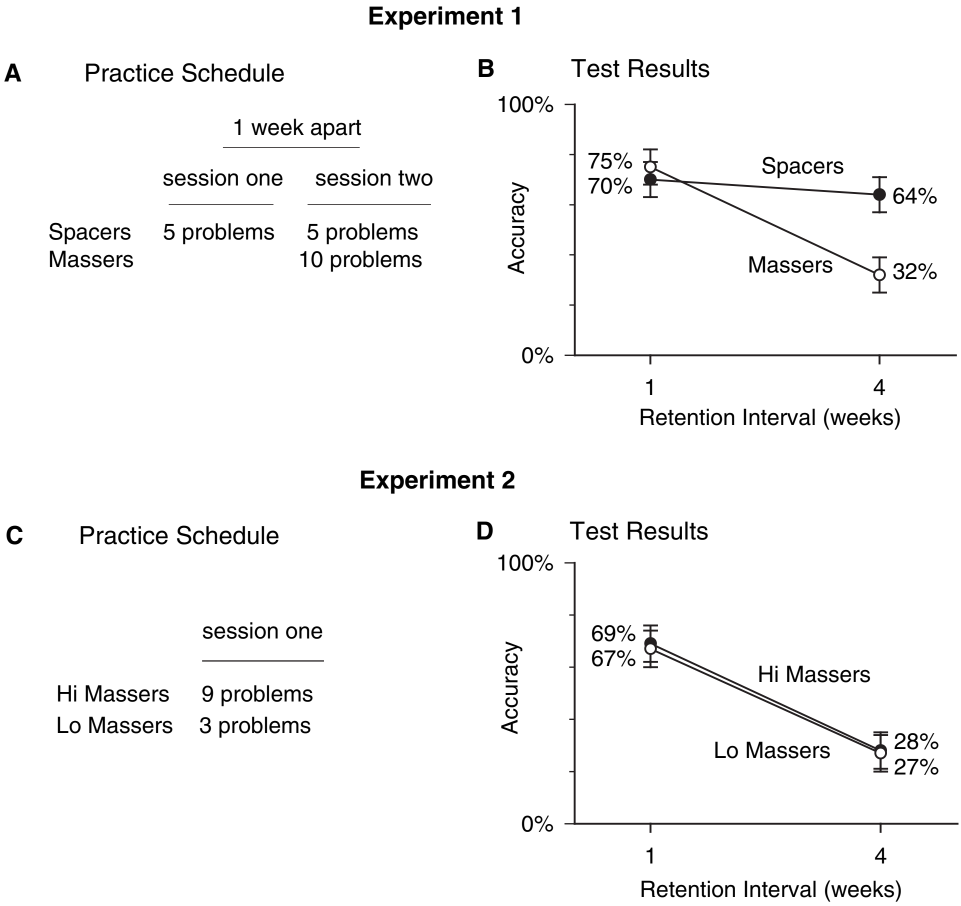

Figure 1. (A) Practise Schedule for Experiment 1. (B) Test Results for Experiment 1. Error bars reflect plus or minus one standard error. (C) Practise Schedule for Experiment 2. (D) Test Results for Experiment 2. Error bars reflect plus or minus one standard error.

EXPERIMENT 2

The second experiment assessed the effect of overlearning on retention by varying the number of practise problems within a single session. The Hi Massers attempted nine practise problems, whereas the Lo Massers attempted only three practise problems, as detailed in Figure 1C. Because the Hi Massers were able to master the task during their first three problems, their additional six practise problems constituted an overlearning strategy. As detailed in the introduction, many researchers have found benefits of overlearning on a subsequent test, but the majority of these experiments relied on single, relatively brief RIs. In this experiment, students were tested either 1 or 4 weeks later.

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

Each student was randomly assigned to one of four groups: Hi Massers with 1-week RI ($n = 28$), Hi Massers with 4-week RI ($n = 22$), Lo Massers with 1-week RI ($n = 24$) and Lo Massers with 4-week RI ($n = 26$).

Procedure

Each student attended two sessions separated by 1 or 4 weeks. At the beginning of the first session, each student was randomly assigned to one of the four conditions. Each student then observed a tutorial consisting of screen projections that included the solutions to Problems 10, 11 and 12 of the Appendix, in that order. As in Experiment 1, these solutions included the two steps illustrated in the Appendix.

Immediately after this tutorial, all students began the practise problems. Each Hi Masser was given 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 difficulty of practise 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 a corresponding number of Lo Massers, thereby equating the difficulty of the first three problems. The remainder of the procedure, including the five test problems and the presentation of feedback after each attempt, was the same as in Experiment 1.

Results and discussion

Practise problems

The tutorial was again sufficient to produce learning, as demonstrated by performance on the first three practise problems (which were the only three practise problems given to the Lo Massers). Of the 100 students, 65 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 practise 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 1- or 4-week test. Consequently, an analysis of variance revealed no main effect of the number of practise problems ($F < 1$) and no two-way interaction ($F < 1$). Yet the main effect of RI was significant, $F\ (1,\ 84) = 33.16$, $p < 0.001$, $\eta_p^2 = 0.28$.

In brief, the threefold increase in the number of same-session practise problems had no observable effect on subsequent test scores at either RI. This null effect of overlearning is not well explained by a lapse in attention by the Hi Massers during their additional six practise problems because these problems were solved with 95% accuracy.

Thus, as more fully detailed in the General Discussion, these results provide no support for the widespread belief 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 present results suggest that the long-term retention of mathematics is aided by distributed practise but not by an overlearning strategy. Specifically, Experiment 1 revealed a boost in long-term retention when practise problems were distributed rather than massed, and Experiment 2 found no benefit of overlearning via a threefold increase in total practise problems. Thus, distributed practise produced a large benefit without requiring extra practise, and overlearning required extra practise yet produced no observable benefit.

As detailed in the introduction, we know of no previously published, unconfounded experiment showing a benefit of distributed practise for a mathematics task other than those requiring only verbatim recall (e.g. $8 \times 5 = 40$). The results of Experiment 1, however, suggest that the superiority of distributed practise over massed practise does, in fact, extend to mathematics tasks that are at least moderately conceptual. Of course, the permutation task was not as conceptually difficult as some mathematics tasks, as every problem was based on the same procedure (unlike a series of mathematical proofs, for instance). Thus, while it remains to be shown whether the current findings extend to far more demanding types of problems, the present results nevertheless lead us to suggest that mathematics students should increase their reliance on distributed practise.

By contrast, the null effect of overlearning in Experiment 2 is obviously at odds with the widespread support for overlearning among educators and researchers. For example, Jahnke and Nowaczyk (1998) advised, 'Practise should proceed well beyond that minimally necessary for an immediate, correct first reproduction' (p. 181). Fitts (1965) concluded that 'The importance of continuing practise beyond the point in time where some (often arbitrary) criterion is reached cannot be overemphasised' (p. 195). Hall (1989) wrote, 'The overlearning effect would appear to have considerable practical value since continued practise 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).

This advocacy for overlearning may reflect previous experimental findings showing benefits of overlearning. Yet, as noted in the introduction, most previous overlearning experiments relied on RIs of 1 week or less (and often less than 1 hour). Thus, the use of moderately long RIs in the present experiment may partly explain the absence of any observed benefit of overlearning.

It is important to note, however, that the null effect of overlearning observed here does not rule out the possibility that a small amount of overlearning may be useful to mathematics students. Strictly speaking, overlearning occurs if a student correctly solves just two problems of the same type (in immediate succession) because the second problem constitutes overlearning. Therefore, even the *Lo* Massers in Experiment 2 relied on a small amount of overlearning because most of them correctly answered two or all three of their practise problems. Consequently, we do not endorse the extreme view that students work only one problem of each type in a given session. Instead, we suggest that assignments should err slightly in the direction of too much practise, perhaps by including three or four problems relating to each new concept in the most recent lesson (in addition to any examples given in the written lesson or class lecture). However, beyond these first three or four problems, the present data suggest that the completion of additional problems of the same type is a terribly inefficient use of study time. Instead, our findings suggest that the student should devote the remainder of the practise session to problems drawn from earlier lessons in order to reap the benefits of distributed practise.

Conceptually, the minimal effect of overlearning on retention can be interpreted as an instance of diminishing returns. That is, with each additional amount of practise 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 practise problems might yield a large increase in a subsequent test score, but each additional practise problem would provide little or no gain unless it is delayed until a later session.

The theoretical question, then, is why these additional practise problems are beneficial only when delayed until a later session. One possibility is that, after a certain amount of practise is completed, additional practise is beneficial only if the initial practise has undergone consolidation, a process by which memory traces are strengthened during the time period immediately after the learning episode. That such consolidation occurs is evidenced by the fact that activities undertaken during this post-learning time period are known to affect subsequent test performance (e.g. Wixted, 2004). For example, retention is increased if learning is followed immediately by sleep (e.g. Plihal & Born, 1997), presumably because sleep enhances consolidation. It may be that consolidation reflects the neural process of long-term potentiation, which requires more than an hour by even the most conservative estimates (e.g. Lu, Kandel, & Hawkins, 1999).

IMPLICATIONS FOR MATHEMATICS TEXTBOOKS

Many mathematics textbooks rely on a format that fosters both overlearning and massed practise. 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 induces overlearning because each assignment includes many problems of the same kind, and it also emphasises massed practise 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, these practise sets emphasise overlearning and massing.

Fortunately, there is an alternative format that minimises overlearning and massed practise while emphasising distributed practise, and it does not require an increase in either the number of practise sets or the number of problems per practise set. With this distributed practise format, each lesson is followed by the usual number of practise problems, but only a few of these problems relate to the immediately preceding lesson. Additional problems of the same type might also appear once or twice in each of the next dozen assignments and once again after every fifth or tenth assignment thereafter. In brief, the number of practise 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.

In addition to its apparent effect on retention, a distributed practise format may also facilitate learning by allowing students ample time to master the particular procedure. For instance, if a student is unable to solve a problem in the assignment immediately after the lesson relating to this type of problem, the student can seek help during the next class meeting (or observe the solution of that problem by the instructor or fellow students) before attempting additional practise problems in the next assignment. The same benefit arises for students who are absent from class when a topic is first presented.

Because a distributed practise format ensures that practise sets include a variety of problem types drawn from previous lessons, instructors who omit lessons need to ensure that each assignment excludes problems corresponding to these omitted lessons. To facilitate this task, the textbook can include an index listing the lesson corresponding to each practise problem so that teachers can more easily avoid assigning problems that students have not learned how to solve. This index would also direct students to the appropriate lesson for help if they are unable to solve a practise problem.

The variety of problems within each practise set that inherently arises with a distributed practise format might reduce the monotony of the assignment while also presenting an additional challenge. Specifically, whereas a massed practise format allows students to solve a problem and then merely repeat the same procedure thereafter, a distributed practise format requires students to recall a variety of procedures from both the immediate and distant past. This added challenge is worthwhile, though, if it boosts retention, thereby providing an instance of what Bjork and his associates have dubbed a 'desirable difficulty' (Christina & Bjork, 1991; Schmidt & Bjork, 1992).

Notably, this mixture of problem types (that arises naturally with the use of a distributed practise format) may provide benefits above and beyond the benefit of distributed practise *per se*. This is because a practise set with a mixture of problem types requires that students learn not only *how* to perform a procedure, but also *which* procedure is needed. For example, the solution of the equation, $x^3 - x = 0$, requires the realisation that factoring is necessary, which allows the equation to be rewritten as, $x\ (x^2 - 1) = 0$, and then, $x(x + 1)(x - 1) = 0$. By contrast, the equation, $x^2 - x - 1 = 0$, cannot be solved by factoring and instead requires the quadratic formula. However, learning to pair the appropriate procedure with each kind of problem is a skill that is not learned during massed practise. For example, if a lesson on the quadratic formula is followed immediately by a dozen problems requiring the quadratic formula, it is obvious to the student that each problem requires the quadratic formula. Of course, this cue is absent when such a problem appears in a cumulative final or arises in a real world application, leaving the student unable to recall which procedure is appropriate. By contrast, in order to complete a practise set with a mixture of problem types, students must know *which* procedure is appropriate and *how* the procedure is performed. Fortunately, mixed practise is an inherent feature of the distributed practise format.

A distributed practise format is used in the Saxon (1997) series of mathematics textbooks, although we are not aware of any published, controlled experiments that compare a Saxon textbook to one with a massed practise format. However, such an experiment may not be particularly informative because a Saxon textbook and a non-Saxon textbook would have numerous differences that would confound the experiment. For example, if such an experiment revealed that one textbook was superior to another, the observed difference might reflect differences in the lessons rather than differences in practise sets. (Neither author is affiliated with a publishing company, although the first author is a former mathematics teacher who used both Saxon and non-Saxon mathematics textbooks.)

A more informative experiment would compare two groups of students who underwent instructional programs that differed only in the temporal distribution of practise problems. That is, each group would complete the same class activities and receive a course packet with the same lessons in the same order. Likewise, the packets would include the same number of practise sets and the same practise problems, but the practise sets would rely on either a distributed or massed practise format.

Textbook publishers could adopt a distributed-practise format with little trouble or cost. For textbooks already in print, this could be accomplished by merely rearranging the practise problems in the next edition, without necessarily altering the lessons. Oddly, practise problems typically receive relatively little attention from publishers and textbook authors, and the practise problems are often written by sub-contracted teachers. Yet the practise sets are just as important as the lessons. In fact, as many mathematics teachers will attest, a majority of their students never read the lessons and instead rely solely on classroom examples before beginning the practise set.

Finally, the benefits of a distributed practise format are equally applicable to computer-aided instruction packages. Unlike textbooks, these algorithms can provide individualised 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 most available computer-aided instruction programs are designed to foster learning rather than retention. These computer programs are easily adapted to incorporate distributed rather than massed practise, and students' compliance to a distributed practise schedule can be verified by ensuring that the program records the date of each use.

CONCLUSION

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

A boost in students' mathematics retention would presumably contribute to an improvement in mathematics achievement, and there is little doubt that there is widespread need for such improvement. 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 practise schedules that proved to be the worst strategies in the experiments reported here.

ACKNOWLEDGEMENTS

We thank the Institute of Education Sciences, U.S. Department of Education, for its generous support (Grants R305H020061, R305H040108). We also thank Christina

Coghlan, Tammy Logan, Kristina Martinez, Erica Porch and Courtnay Rosene for their assistance with data collection.

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.

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

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

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

Bjork, R. A. (1988). Retrieval practise 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 practise 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 practise. *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 practise 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 practise 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 practise 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.
Lu, Y. F., Kandel, E. R., & Hawkins, R. D. (1999). Nitric oxide signaling contributes to late-phase LTP and CREB phosphorylation in the hippocampus. *Journal of Neuroscience, 19*, 10250–10261.
Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience, 9*, 534–547.
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 practise 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 practise: 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 practise: 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.
Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235–269.

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. *aabbbb*). For n items and k unique items, the number of permutations equals $n!/(n_1!\ n_2!\ \ldots\ n_k!)$, where $n_i =$ number of repetitions of item i. Thus, for *aabbbb*, the number of permutations equals 6!/(2! 4!), or 15. However, students were not shown this formula because we expected that the participants would have vastly different amounts of experience with factorial notation.

Students were instead taught a two-step procedure that excluded factorials and variables. For example, for the sequence *aabbbb* (with six letters, two occurrences of *a*, and four occurrences of *b*), the solution began with the fractional expression,

$$\frac{6 \bullet 5 \bullet 4 \bullet 3 \bullet 2}{(2)(4 \bullet 3 \bullet 2)}.$$

This expression was supplemented by comments explaining the choice of the first digit of each factorial expression. Thus, in this example, 'six letters' appeared to the left of the numerator, and, just below the first digit of each parenthetical expression within the denominator, there appeared the comments '*a* appears two times' and '*b* appears four times'. The second step of the solution required simplification by cancellation of like factors from the numerator and denominator, followed by arithmetic. For this example, students saw

$$\frac{6 \bullet 5}{2} = 15.$$

PRACTISE SESSION PROBLEMS

The practise problems are listed below. In Experiment 1, the tutorial included Problems 11 and 12, and the 10 practise problems were Problems 1 through 10, in that order. In Experiment 2, the tutorial included Problems 10, 11 and 12. The Hi Massers worked Problems 1 through 9, and the Lo Massers worked just three of these nine problems. The order of presentation varied for both groups, as explained in the Procedure section of Experiment 2.

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. aabbbbbb	28
11. abbcccc	105
12. abcccccc	56

TEST PROBLEMS

In both experiments, the test included the following five problems in the order shown.

abcc	12
aabbb	10
aabbcc	90
aaabbbb	35
aaaabbbb	70

