APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **24**: 837–848 (2010)
Published online 30 July 2009 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/acp.1598

The Effects of Interleaved Practice

KELLI TAYLOR and DOUG ROHRER*

Department of Psychology, University of South Florida, USA

SUMMARY

Previous research shows that interleaving rather than blocking practice of different skills (e.g. *abcbcacab* instead of *aaabbbccc*) usually improves subsequent test performance. Yet interleaving, but not blocking, ensures that practice of any particular skill is distributed, or *spaced*, because any two opportunities to practice the same task are not consecutive. Hence, because spaced practice typically improves test performance, the previously observed test benefits of interleaving may be due to spacing rather than interleaving *per se*. In the experiment reported herein, children practiced four kinds of mathematics problems in an order that was interleaved or blocked, and the degree of spacing was fixed. The interleaving of practice impaired practice session performance yet doubled scores on a test given one day later. An analysis of the errors suggested that interleaving boosted test scores by improving participants' ability to pair each problem with the appropriate procedure. Copyright © 2009 John Wiley & Sons, Ltd.

When practice is *interleaved* rather than *blocked*, the practice of different skills is intermixed rather than grouped by type (e.g. *abcbcacab* instead of *aaabbbccc*). In most introductory statistics textbooks, for example, the practice problems requiring a particular statistical test (e.g. independent samples t test) are blocked together in a set that typically follows the lesson on that statistical test. With interleaved practice, by contrast, each lesson is followed by a set of practice problems drawn from many previous lessons so that no two problems of the same kind appear consecutively, thereby requiring students to choose the appropriate statistical test on the basis of the problem itself. Several studies have shown that interleaving yields better subsequent test performance than does blocking. In these studies, however, practice in the interleaving condition was distributed or *spaced*, and, as detailed below, this confound worked in favour of the observed benefits of interleaving. The present study assessed the effects of interleaving while equating the degree of spacing.

Literature

Most of the previous studies of interleaving examined its effects on motor skills (e.g. Carson & Wiegand, 1979; Hall, Domingues, & Cavazos, 1994; Keller, Li, Weiss, & Relyea, 2006; Landin, Hebert, & Fairweather, 1993; Shea & Morgan, 1979). For instance, in the study reported by Hall et al., college baseball players practiced hitting three types of pitches (e.g. curve ball) that were either blocked by type or systematically interleaved. During a

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

subsequent test in which the three types of pitches were interleaved (as in an actual game), hitting performance was greater if practice had been interleaved rather than blocked. A similar benefit was observed in a study of basketball shooting by Landin et al. (1993).

Fewer studies have examined how interleaving affects the learning of cognitive skills, but these studies have nevertheless consistently demonstrated that the interleaving of practice boosts subsequent test performance. For example, in a study reported by Kornell and Bjork (2008), participants saw numerous paintings by each of a dozen artists during the learning phase, with each painting appearing with the artist's name. The order in which the paintings appeared was manipulated so that paintings were either blocked by artist or interleaved across artists. On a subsequent test in which participants saw *novel* paintings by these same artists, participants were better able to identify the artist of each painting if the learning order had been interleaved rather than blocked.

In an interleaving study by Rohrer and Taylor (2007), college students learned how to find the volume of four obscure solids by completing eight practice problems for each of the four solids, and the practice problems were either blocked by type or interleaved. One week after the final practice problem, participants returned for a test consisting of novel problems. Interleaving of practice tripled test scores (63% *vs.* 20%), Cohen's $d = 1.34$. (However, interleaving sharply *impaired* accuracy during practice, 60% *vs.* 89%, Cohen's $d = 1.06$, and this finding, which others have reported as well, is revisited in the Discussion.)

Most recently, Le Blanc and Simon (2008) assessed the effects of interleaving different kinds of mathematics practice problems on both test performance and participants' ability to estimate their test performance. Participants learned to find the volume of four solids—the same as those used in Rohrer and Taylor (2007)—and, once again, interleaving of practice dramatically boosted scores on a test given one week after practice. Furthermore, interleaving of practice improved participants' self-reported estimates of their test scores, regardless of whether this judgment of learning (JOL) was made immediately after practice or immediately after the test.

A confound

Although previous studies have revealed a consistently large test benefit of interleaving, the design of these studies clouds the interpretation of what exactly is producing this benefit. This is illustrated by the following procedure from a prototypical interleaving study in which every participant practices four different kinds of skills (*a*, *b*, *c* and *d*) an equal number of times but in a different order:

Interleaved :	a b c d	b a d c	d c b a	c d a b
Blocked :	a a a a	b b b b	c c c c	d d d d

Thus, interleaved practice ensures that the amount of practice devoted to a particular skill is spaced or distributed across multiple episodes (and separated by intervening tasks), whereas blocked practice guarantees that practice of any particular skill is concentrated or *massed* into an uninterrupted period of time. Consequently, the previously observed benefits of interleaving might reflect, partly or entirely, the contribution of spacing. This confound is problematic because spacing usually boosts subsequent test performance—indeed, this *spacing effect* is one of the most robust effects in the area of learning (for reviews, see, for example, Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1987; Bjork, 1994; Donovan & Radosevich, 1999; Rohrer, 2009; for experimental papers, see, for

DOI: 10.1002/acp

example, Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E., 1993; Bloom & Shuell, 1981; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Cook, 1934; Kornell & Bjork, 2008; Lee & Genovese, 1988; Mumford et al., 1994; Pyc & Rawson, 2007; Rea & Modigliani, 1985; Rohrer & Taylor, 2006; Seabrook, Brown, & Solity, 2005).

We are not, however, suggesting that previous interleaving studies are flawed because these studies conflated the effects of interleaving and spacing. (In fact, as summarized above, one of these studies is ours.) Rather, we assert that previous interleaving studies are informative, because, in real-world scenarios, interleaved practice is usually spaced while blocked practice is usually massed. Thus, the results of previous interleaving studies provide a measure of the benefits that can be realized if blocked, massed practice is replaced by interleaved, spaced practice. We also should emphasize that the authors of most if not all of the previous interleaving studies fully realized that their studies conflated interleaving and spacing; in some cases, in fact, the interleaving manipulation is explicitly described as one that combines interleaving and spacing (e.g. Kornell & Bjork, 2008).

Nevertheless, it is important to measure the effects of interleaving while holding fixed the degree of spacing in order to assess the independent effects of interleaving and better understand exactly how interleaving affects learning. Yet, whereas hundreds of studies have assessed the spacing effect, the effects of interleaving *per se* have not, to our knowledge, been previously assessed. Therefore, we designed the present study to isolate the effect of interleaving by equating the degree of spacing in the interleaved and blocked practice conditions.

METHOD

Participants

Two sessions were completed by 24 children (12 boys and 12 girls) who were 10 or 11 years of age ($M = 10.5$, $SD = 0.51$). Each was a fourth grade student at an elementary school in St. Petersburg, Florida. An additional two students completed the first session but failed to return for the second session (because they were absent from school), and their data were excluded from all analyses.

Task

Participants were taught to solve four kinds of mathematics problems, each of which concerned one of the following four parts of a prism: face, corner, edge or angle (which is defined as the interior angle of a face), as shown in Figure 1. In each problem, participants were given the number of base sides (b), the number of sides of the prism base, and were required to find the total number of faces, corners, edges, or angles by using one of the formulas shown in Figure 1. The value of b ranged between 3 and 30, and no particular b value was paired with the same prism part more than once during the entire study, which is to say that no problem appeared more than once. Participants were taught a two-step solution, as detailed below in the Procedure section.

Procedure

Participants attended a *practice session* and a *test session* one day later. Each participant was randomly assigned to either the interleaved or blocked group prior to the practice

DOI: 10.1002/acp

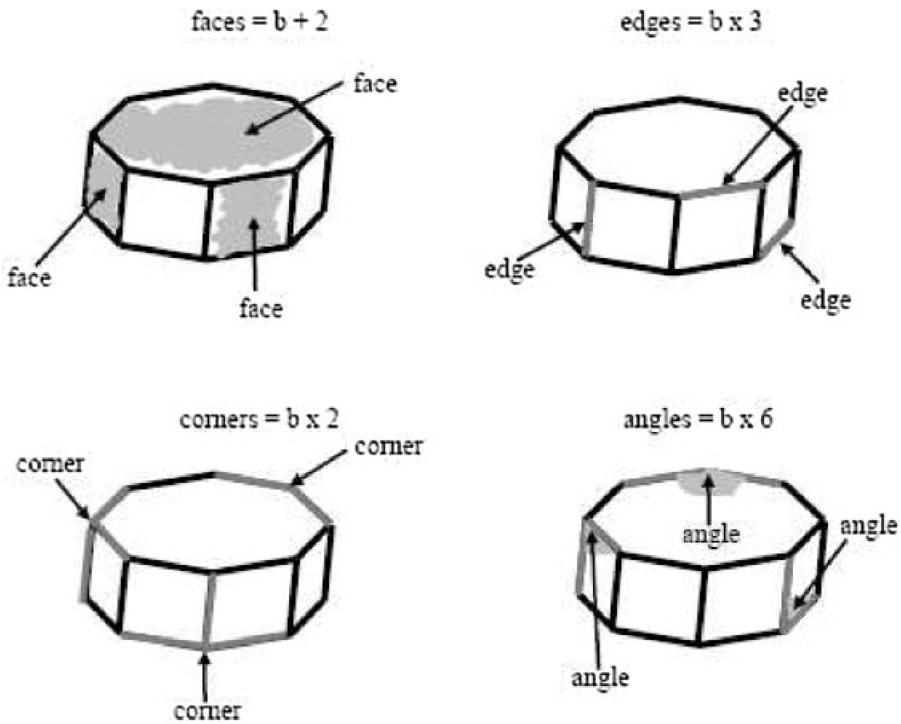

Figure 1. *Formulas and tutorial pictures*. The illustrations above were presented only during the tutorial. For problems seen during the test or the full-problem phase within the practice session, participants were given the number of base sides (*b*) and required to determine the number of faces, edges, corners, or angles (the latter was defined to include every interior angle of every face). See text for details

session. Each group saw the same tutorials, examples, practice problems, and test problems, although the order of the examples and practice problems depended on whether practice was interleaved or blocked, as detailed below.

During the practice session, participants observed a computer-generated visual presentation and wrote their responses in individual booklets. Participants first completed a pre-test consisting of one of each of the four kinds of problems (i.e. face, corner, edge and angle). Although participants were not required to show their solution on the pre-test, none provided a single correct answer. The pre-test was followed immediately by a tutorial that included illustrations and definitions of a prism and the parts of a prism relevant to the task: base sides, faces, edges, corners and angles (Figure 1). The remainder of the practice session consisted of two phases.

Partial-problem phase

Because pilot studies with fifth grade students from another school revealed that most of the students were overwhelmed when the initial practice problems required the entire solution, we designed the present study so that the first eight practice problems of each kind (P1–P8, where *P* connotes *Partial*) required only the first step of the solution. Specifically, the term for one of the four prism features (e.g. *faces*) appeared onscreen (5 seconds), which prompted participants to write the appropriate expression ($b+2$) in their booklet. Immediately afterward, the correct formula (faces $= b+2$) appeared onscreen (5 seconds), and participants were required to mark their answer as correct or incorrect to ensure that they were aware of their errors. In addition to the four types of problems (*a*, *b*, *c* and *d*), participants saw one solved example (7 seconds) of each type (*A*, *B*, *C* and *D*), and they completed 30-s filler tasks (*f*) that were comprised of brief puzzles unrelated to geometry or algebraic substitution.

The problems, examples, and filler tasks during the partial-problem phase were ordered as follows:

Interleaved : ABCD abcd cadb dcba bdac bdac dcba cadb abcd ffffffffffffffffffffffffff
Blocked : Aafafafafafafafa Bbfbfbfbfbfbfbfb Ccfcfcfcfcfcfcfc Ddfdfdfdfdfdfdfd

Thus, with interleaved practice, participants viewed all four examples before the first practice problem, and the ensuing 32 problems (8 problems of each kind) were intermixed so that three constraints were satisfied. First, each set of four consecutive problems (e.g. 1–4, 5–8, etc.) included one problem of each kind. Second, any two problems of the same kind (e.g. 'edge problems') were separated by at least one problem of another kind. Third, each problem and the next occurrence of the same kind of problem were separated by an average of exactly three intervening problems. For example, the eight 'corner problems' (represented by *c* above) were separated by 1, 4, 5, 3, 1, 2 and 5 intervening problems, respectively, for an average of 3 intervening problems, which, at 10 seconds per problem, translates to an average spacing gap of 30 seconds. Immediately after completing these problems, the interleaved group completed every filler task. With blocked practice, all eight problems of each kind (e.g. the 'edge problems') appeared immediately after the corresponding example. For instance, the 'edge example' was followed immediately by all eight 'edge problems'. However, no two problems appeared in immediate succession, as a 30-s filler task preceded all but the first problem of each kind, thereby ensuring that a task-filled spacing gap between each problem and the next problem of the same kind was 30 seconds in duration, which equalled the average spacing gap provided by interleaved practice.

Full-problem phase

In the final phase of the practice session, problems required participants to perform both steps of the solution: (1) recall and write the appropriate formula, and (2) substitute the appropriate value for *b*. For example, the question 'A prism with 17 base sides has how many edges'? would require participants to write '$b \times 3$' and then '17×3'. They were not, however, asked to simplify the latter expression by performing the necessary addition or multiplication (e.g. $17 \times 3 = 51$), because arithmetic was a skill they had learned prior to the study. Each problem appeared on its own page within the booklet, and a visual cue and auditory tone from the computer prompted participants to turn the page and attempt to write both steps of the solution within 15 seconds. Immediately afterward, both steps of the solution appeared onscreen (10 seconds), prompting participants to check both steps of their solutions and mark each step as correct or incorrect. In addition to the four types of problems (*a*, *b*, *c* and *d*), participants saw one solved example (25 seconds) of each type (*A*, *B*, *C* and *D*) and completed 75-s filler tasks (*f*) like those in the first phase.

DOI: 10.1002/acp

The problems, examples, and filler tasks in the full-problem phase were ordered as follows:

Interleaved : ABCD abcd bdac abcd fffffff
Blocked : Aafafa Bbfbfb Ccfcfc Ddfdfd

Thus, interleaved practice began with all four examples, and the order of the 12 practice problems (3 of each kind) was subject to the same constraints governing the partial-problem phase. First, each set of four problems (i.e. 1–4, 5–8, etc.) included one problem of each kind. Second, no two problems of the same kind appeared consecutively. Third, each problem and the next occurring problem of the same kind were separated by an average of exactly three intervening problems, which translates to an average spacing gap of 75 seconds. The interleaved group then completed every filler task. The blocked practice group encountered every practice problem of each particular kind (e.g. all three 'edge problems') immediately after they saw the corresponding example. However, each pair of consecutive problems of the same kind (e.g. two 'edge problems') were separated by a 75-s filler task, so that the duration of the spacing gap between two consecutive problems of the same kind equalled that provided by interleaving.

Tests

One day after the practice session, participants returned for two paper-and-pencil tests. None of the problems from the practice session appeared on either test, and feedback was not provided. *Test 1* (2 minutes) included four problems—one of each kind (i.e. face, corner, edge and angle)—and each required the full solution (i.e. write the appropriate formula and substitute for b). The immediately following *Test 2* (2 minutes) also consisted of four problems, one of each kind, but, unlike Test 1, all four formulas (e.g. faces $= b + 2$) were provided at the top of the page so that participants did not need to recall which expression (e.g. $b + 2$) corresponded to each prism feature (e.g. face).

RESULTS

Practice problems

Figure 2A shows the mean accuracy for the eight problems of each kind in the partial-problem phase (P1–P8) and the three problems of each kind in the full-problem phase (F1–F3). Across all problems in the partial-problem phase, mean accuracy was far worse when practice problems were interleaved rather than blocked (68% *vs.* 99%), $t(22) = 4.94$, $p < .01$, Cohen's $d = 2.02$. Likewise, interleaving impaired accuracy during the full-problem phase (79% *vs.* 98%), $t(22) = 2.94$, $p < .05$, Cohen's $d = 1.20$. Even for the final practice problem of each kind (F3), on which the interleaved group achieved their best score in the entire study, the interleaved practice group performed reliably worse than the blocked practice group (81% *vs.*100%), $t(22) = 2.46$, $p < .05$, Cohen's $d = 1.01$. In summary, practice performance was sharply *impaired* by interleaving.

Tests

Mean accuracy on Test 1, which required participants to pair each kind of problem with the appropriate solution (as in the practice session), was more than doubled by the interleaving of practice (77% *vs.* 38%), $t(22) = 2.96$, $p < .01$, Cohen's $d = 1.21$. Thus, as shown in

DOI: 10.1002/acp

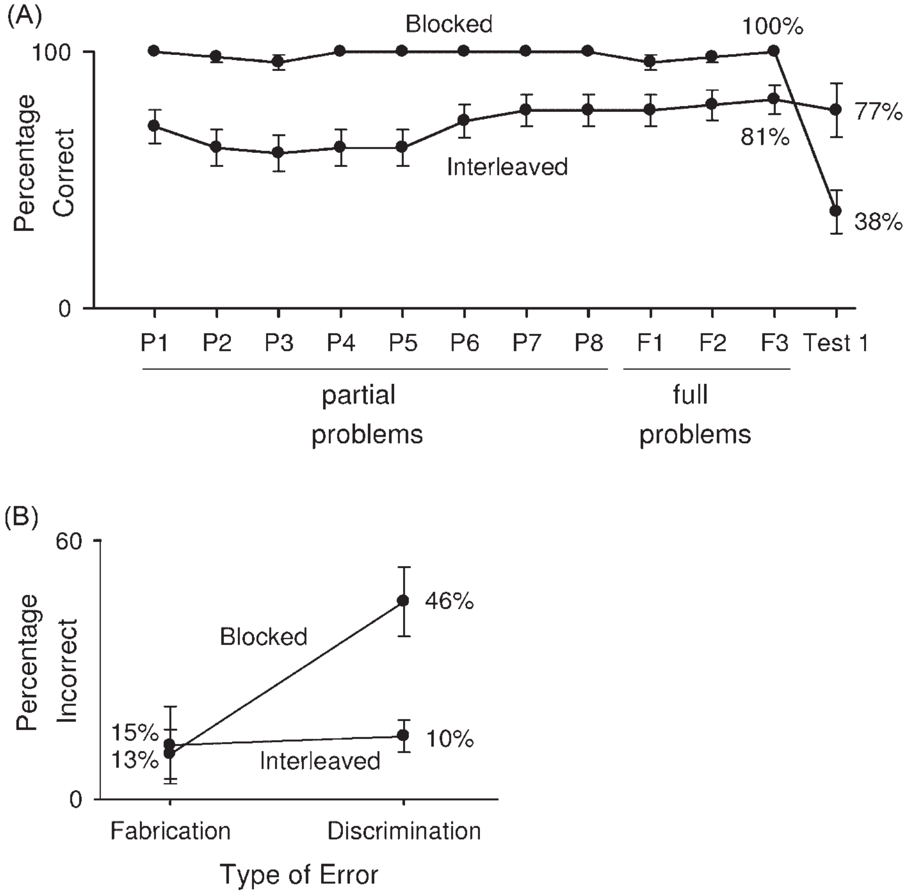

Figure 2. *Results*. (A) Percentage correct for Practice Problems and Test 1. Data averaged across the four kinds of problems. The partial-problem phase (P1–P8) required only the first step of the solution, whereas the full-problem phase (F1–F3) required both steps. Error bars equal plus or minus one standard error. (B) Test errors. Data points represent the percentage of Test 1 problems that led to either a discrimination or fabrication error (see text for more detail). Error bars equal plus or minus one standard error

Figure 2A, interleaving of practice impaired practice performance yet improved Test 1 performance. This interaction between strategy (interleaving *vs.* blocked) and session (practice *vs.* Test 1), where practice performance is measured by the full-problem phase (F1–F3), was shown by a two-way analysis of variance to be significant, $F\,(1, 22) = 29.33$, $p < .001$, $\eta_p^2 = .57$. Less importantly, and not surprisingly, this analysis of variance also revealed a main effect of session, with practice session accuracy exceeding Test 1 accuracy, presumably because instruction and feedback were provided during the practice session but not during the test session, $F\,(1,\,22) = 33.67$, $p < .001$, $\eta_p^2 = .60$.

Whereas interleaved practice improved performance on Test 1, which required participants to both choose the appropriate formula and perform the procedure (steps 1 and 2), interleaving provided a small, unreliable benefit on Test 2, which required participants to only *perform* the procedure (100% *vs.* 90%), $t(22) = 1.19$. Admittedly, though, this null finding might have reflected a ceiling effect.

DOI: 10.1002/acp

Test errors

With the exception of a single problem that was left unsolved by a single participant, every error on Test 1 was one of two types. With a *discrimination error*, the participant mistakenly used one of the other three formulas they had encountered during the practice session. With a *fabrication error*, the participant used a formula (veridical or otherwise) that they had not encountered during the experiment.

The likelihood of each kind of error occurring on Test 1 is shown in Figure 2B. As shown, blocked practice and interleaved practice produced about the same number of fabrication errors (15% *vs.* 13%, respectively), but interleaving of practice dramatically reduced the frequency of discrimination errors (46% *vs.* 10%). This interaction between strategy (interleaved *vs.* blocked) and type of error (fabrication *vs.* discrimination) was confirmed by a two-way analysis of variance, $F\,(1, 22) = 5.70$, $p < .05$, $\eta_p^2 = .21$, and *post hoc* Tukey tests revealed that discrimination errors, but not fabrication errors, were more common when practice was blocked rather than interleaved, $p < .001$. (Of lesser relevance, this analysis of variance also revealed two main effects, both $p\text{s} < .05$). That blocking increased the number of discrimination errors without detectably affecting the number of fabrication errors is consistent with the hypothesis that interleaved practice boosted test scores by improving participants' ability to pair each problem with the appropriate procedure. This hypothesis is revisited further below.

DISCUSSION

In the study reported here, the interleaving of different kinds of practice problems more than doubled subsequent test scores (77% *vs.* 38% on Test 1) even though the degree of spacing was held constant. To our knowledge, this is the first study to assess the benefit of interleaving while controlling for the benefit of spacing, as previous studies have assessed the conjoint effects of interleaving and spacing. Thus, although the spacing effect is one of the most robust effects in learning, the present data demonstrate that interleaving *per se* can improve test scores by a large margin.

Whereas the interleaving of practice boosted *test* performance, it impaired performance during the *practice* session (Figure 2A). Hence, interleaving proved to be an instance of what Robert Bjork and his colleagues have termed a *desirable difficulty*, because the difficulty incurred during the practice session proves to be ultimately worthwhile (e.g. Bjork, 1994; Christina & Bjork, 1991; Schmidt & Bjork, 1992). These authors have previously identified both interleaving and spacing as desirable difficulties.

Caveats and limitations

Although the effects of interleaving reported here were large, it remains unknown whether these effects will generalize to other kinds of materials and procedures. For instance, the delay between the practice session and the test was only one day, far shorter than the retention typically sought by learners. In addition, performance by the blocked practice group was nearly perfect throughout most of the practice session (Figure 2A), and this ceiling effect raises the possibility that a procedure with fewer practice problems might have produced a smaller test benefit of interleaving. That is, if blocked practice, but not interleaved practice, provides sharply diminishing benefits with each successive problem

DOI: 10.1002/acp

(as measured by a subsequent test), the concomitant reduction in test scores will be greater for the blocked practice group than for the interleaved practice group, thereby reducing the difference between the two groups. In brief, the test benefit of interleaving may require a sufficiently large amount of practice.

Furthermore, the benefits of interleaving practice likely depend on the similarity of the tasks. In the present study, the different kinds of problems required similar solutions, and this similarity presumably contributed to the observed effects. Hence, if several tasks are instead *easily* distinguished from each other, interleaving might be less beneficial. Indeed, interleaving might exert its benefit by improving discriminability, as detailed below.

A final caveat is that task discriminability can prove difficult in different ways, and each might not afford the benefit of interleaving reported here. In the present study, for instance, participants struggled to pair each kind of problem with the appropriate solution, but it was easy to identify the *kind* of problem (because the statement of each problem included a term such as 'faces' or 'edges'). In other scenarios, identifying (or classifying) the problem or task is the greater challenge. For example, different kinds of inferential statistics problems are usually stated in a superficially similar fashion—each includes data and the instruction to assess statistical significance. Thus, learners must recognize subtle features of the problem (e.g. parametric *vs.* non-parametric, repeated measures *vs.* independent measures, etc.) that allow them to identify which *kind* of problem it is (e.g. one requiring a Mann–Whitney U test). In brief, the present study assessed the effects of interleaving with a single set of tasks, and only additional research can determine the extent to which the observed effects will generalize. Still, given the magnitude of the test benefit observed here (Cohen's $d = 1.21$), a smaller effect might still be moderately large.

A theoretical account of the benefit of interleaving

Perhaps the most parsimonious explanation of the test benefit provided by interleaving of practice is that interleaving improves discriminability. That is, because interleaving requires participants to repeatedly switch between different kinds of tasks, they must learn how to pair each kind of task with its appropriate procedure. In the present study, for instance, blocked practice allowed participants to completely ignore whether a problem concerned faces, edges, corners, or angles, because every problem in each block concerned the same feature. By contrast, interleaved practice required participants to learn which formula is appropriate for which prism feature. This explanation of how interleaved practice improves test performance is consistent with the finding observed here that blocked practice was nearly as useful as interleaved practice when the test problem did not require participants to pair each problem with its appropriate procedure (i.e. on Test 2 but not Test 1). This account also predicts that the interleaving of practice should reduce the number of discrimination errors occurring during Test 1 (i.e. errors in which participants choose a procedure that is appropriate for a different kind of task) without affecting the frequency of other kinds of errors, and this finding also was observed (Figure 2B).

Practical implications for mathematics learning

If interleaving does, in fact, improve discrimination, it would seem ideally suited for mathematics learning, because the difficulty of many mathematics problems stems from the challenge of identifying the appropriate procedure. Thus, as with the statistics dilemma posted at the outset of the paper, students need to know not only *how* to perform a particular

DOI: 10.1002/acp

statistical test (e.g. independent samples *t* test) but also *which* statistical test is appropriate for a given problem. Indeed, choosing the appropriate statistical test is arguably the more important of the two skills, as statistical software can perform the statistical test. Of course, choosing the appropriate test is not always easy, because few statistical problems (whether arising during a course or a research project) explicitly indicate which statistical test is appropriate.

When practice problems are blocked, however, students can successfully solve a set of practice problems without learning how to pair a problem with the skill. Indeed, because all of the problems relate to the topic—typically the one presented in the immediately preceding lesson—students can choose the appropriate procedure for each practice problem *before* they read the problem. While this reduces the difficulty of the practice problems, students are effectively relying on a crutch. Unfortunately for students, this weakness is exposed when these same kinds of problems appear on a cumulative exam, standardized test, or during a subsequent research career. By contrast, interleaved practice gives students an opportunity to practice pairing each kind of problem with the appropriate procedure.

Far from being limited to statistics courses, the difficulty of pairing a problem with the appropriate procedure or concept is ubiquitous in mathematics. For example, the notorious difficulty of word problems is due partly to the fact that few word problems explicitly indicate which procedure or concept is appropriate. For example, the word problem, 'If a bug crawls eastward for 8 m and then crawls northward for 15 m, how far is it from its starting point'? requires students to infer the need for the Pythagorean Theorem. However, no such inference is required if the word problem appears immediately after a block of problems that explicitly indicate the need for the Pythagorean theorem (e.g. if the legs of a right triangle have lengths 8 and 15 m, what is the length of its hypotenuse?). Thus, blocked practice can largely reduce the pedagogical value of a word problem.

As a final example, it should be noted that blocked practice may facilitate students' failure to discriminate between different kinds of problems even when these kinds of problems are *not* superficially similar. In elementary school, for example, students are ordinarily taught to find both the *greatest common factor* of two integers and the *least common multiple* of two integers. Thus, the instructions for these two kinds of problems are easily distinguished from each other ('Find the greatest common factor ...' *vs.* 'Find the least common multiple ...'). However, if the practice problems of each kind are blocked, students can ignore the instruction and instead focus solely on the information that varies from problem to problem (i.e. the pair of integers). Students can then solve problems by merely repeatedly performing the same procedure without giving much thought to why it is appropriate.

Despite the potential benefits of interleaving, most mathematics textbooks rely primarily on blocked practice. Moreover, it would be easy for authors of mathematics textbooks (and instructional software) to include more interleaved practice, as this would require a mere rearrangement of practice problems with the next edition. This would allow students to reap the large benefit of spacing, which has been demonstrated many times, as well as the benefit of interleaving, which, in light of the data presented here, might be no less robust.

ACKNOWLEDGEMENTS

This research was completed as part of the first author's doctoral dissertation and was supported by the Institute of Education Sciences in the U.S. Department of Education

(Grants R305H040108 and R305B070537). The opinions expressed are those of the authors and do not necessarily represent the views of the Institute of Education Sciences.

REFERENCES

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

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

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.

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

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *11*, 1095–1102.

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

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

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

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

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills*, *78*, 835–841.

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

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

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

Le Blanc, K., & Simon, D. (2008). Mixed practice enhances retention and JOL accuracy for mathematical skills. *Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL. November, 2008*

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

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

Pyc, M. A., & Rawson, K. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*, 1917–1927.

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.

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, *40*, 4–17.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209–1224. DOI: 10.1002/acp.1266

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

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
Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179–187.

DOI: 10.1002/acp