Educ Psychol Rev (2012) 24:355–367
DOI 10.1007/s10648-012-9201-3

REVIEW ARTICLE

Interleaving Helps Students Distinguish among Similar Concepts

Doug Rohrer

Published online: 28 July 2012

Abstract When students encounter a set of concepts (or terms or principles) that are similar in some way, they often confuse one with another. For instance, they might mistake one word for another word with a similar spelling (e.g., allusion instead of illusion) or choose the wrong strategy for a mathematics problem because it resembles a different kind of problem. By one proposition explored in this review, these kinds of errors occur more frequently when all exposures to one of the concepts are grouped together. For instance, in most middle school science texts, the questions in each assignment are devoted to the same concept, and this *blocking* of exposures ensures that students need not learn to distinguish between two similar concepts. In an alternative approach described in this review, exposures to each concept are *interleaved* with exposures to other concepts, so that a question on one concept is followed by a question on a different concept. In a number of experiments that have compared interleaving and blocking, interleaving produced better scores on final tests of learning. The evidence is limited, though, and ecologically valid studies are needed. Still, a prudent reading of the data suggests that at least a portion of the exposures should be interleaved.

Keywords Interleave · Blocked · Spacing · Math · Learning

In virtually any academic discipline, students encounter each important concept more than once, and creators of textbooks and other kinds of learning materials must therefore decide, incidentally or otherwise, how these exposures will be arranged. For example, a science class assignment might focus on a single principle (e.g., photosynthesis) or cover multiple principles. These kinds of scheduling decisions are often dismissed as inconsequential or even dull, which might explain why this kind of intervention receives little or no attention in textbooks and courses devoted to curriculum and instruction. Yet the timing of learning exposures can dramatically affect learning outcomes. Many experiments have shown that merely rearranging the order in which students encounter examples or questions, without

D. Rohrer (✉)
Psychology, University of South Florida, PCD 4118G, Tampa, FL 33620, USA
e-mail: drohrer@usf.edu

altering the number or nature of these exposures, can boost scores on final tests of learning. This review examines one kind of scheduling intervention that appears to be especially useful when students must learn to distinguish among similar concepts.

Discrimination Learning

When students must learn to distinguish among similar concepts, they often confuse one with the other. For example, biology students are asked to distinguish among the genetic processes of transcription, transduction, transformation, and translation—four terms with similar spellings and meanings. Not surprisingly, a greater degree of similarity makes the task more difficult (e.g., Skinner 1933). A failure to distinguish, or discriminate, between two concepts (or terms or principles or stimuli) is called a *discrimination error*, and learning to make these distinctions is *discrimination learning*.

The genetics example is obviously cherry picked, and that raises a question that underlies the practical significance of this review: how often must students make difficult discriminations? Indeed, most of the research on discrimination learning examines skills or concepts learned *outside* the classroom. For instance, there are many studies of the subtle discriminations made during speech comprehension (e.g., "pa" vs. "ba") and face recognition, but children typically master these skills before they begin school. In other studies of discrimination learning, subjects learned skills needed only by experts, such as distinguishing among different species of birds. This is not to say that discrimination expertise is necessarily impractical—indeed, medical doctors must learn to distinguish among diseases with similar symptoms. Still, the majority of discrimination learning studies employ tasks that few students encounter.

Yet subtle discriminations are required in several academic disciplines. In learning a foreign language, for instance, students encounter pairs of words that are easily confused with each other, as illustrated by Spanish word pairs like pero–perro, arrollo–arroyo, ciento–siento, cerrar–serrar, cima–sima, and halla–haya. Vocabulary confusions also occur in a student's first language, and even English-fluent adults struggle with pairs like affect–effect, allusion–illusion, ascent–assent, and appraise–apprise. Especially fiendish discriminations appear in the sciences because groups of two or more terms often have similar spellings and similar meanings, and examples include mitosis–meiosis, oviparous–ovoviviparous, chromatid–chromatin, phagocytosis–pinocytosis, afferent–efferent, adhesion–cohesion, solution–solute–solvent, ion–anion–cation, and glycerol–glycogen–glucagon.

Furthermore, discrimination learning plays a central role in the mastery of mathematics and certain physical sciences. Proficiency in mathematics is measured *solely* by the ability to solve problems, and this in turn demands that students learn to distinguish between superficially similar kinds of problems requiring different strategies. For instance, because students learn to multiply fractions ($^1/_2 \times {}^1/_3$) by multiplying "tops and bottoms," they sometimes mistakenly add fractions ($^1/_2 + {}^1/_3$) by "adding tops and bottoms" (Siegler 2003). In other words, identifying an appropriate strategy for a problem requires students to identify the category, or kind of problem, to which a novel problem belongs. This kind of categorization is required in physics, too. For example, some incline motion problems are solved with Newton's second law, yet others require an understanding of the law of conservation of energy. In fact, this specific example was used in a classic study reported by Chi *et al.* (1981), in which experts, but not novices, categorized a physics problem on the basis of its underlying principle (Newton's second law or the law of conservation of energy) rather than its superficial features (block sliding down an incline). These issues are revisited in a later section devoted to mathematics learning.

Interleaving

When students must learn similar concepts (or skills or terms or principles), the exposures to these concepts are typically arranged in one of two ways. Most often, exposures to each concept are grouped together. For instance, a physical science unit on gravity might include three assignments: one with three questions on pendulums ($a_1a_2a_3$), one with three questions on free fall ($b_1b_2b_3$), and one with three questions on incline motion ($c_1c_2c_3$). This means that the exposures are *blocked* by concept ($a_1a_2a_3b_1b_2b_3c_1c_2c_3$). In an alternative approach described in this review, exposures to concepts are *interleaved* ($a_1b_1c_1b_2c_2a_2c_3b_3a_3$) so that a question on one concept is followed by a question on a different concept. Readers who are familiar with the intervention of spacing might wonder whether interleaving is another term for spacing, but the two interventions are distinct, as detailed further below. Also, though much of this review is devoted to comparisons of interleaving and blocking, students could rely on a combination of the two strategies. For example, initial exposures to a concept could be blocked, and subsequent exposures could appear within interleaved assignments. This hybrid approach is considered later in this review.

The prototypical interleaving experiment is illustrated in Fig. 1a. Problems or questions on each one of several concepts are interleaved or blocked during the learning phase, and students are later tested. Interleaving (rather than blocking) typically improves final test scores, and this benefit is defined here as the *interleaving effect*. A simple example of an interleaving study, though it involved motor skills rather than cognitive skills, was reported by Hall *et al.* (1994). They asked college baseball players to practice hitting three types of pitches (fastball, curveball, and change-up), and the 45 practice pitches were either blocked by type (15 fastballs, 15 curveballs, and 15 change-ups) or interleaved (fastball, curveball,

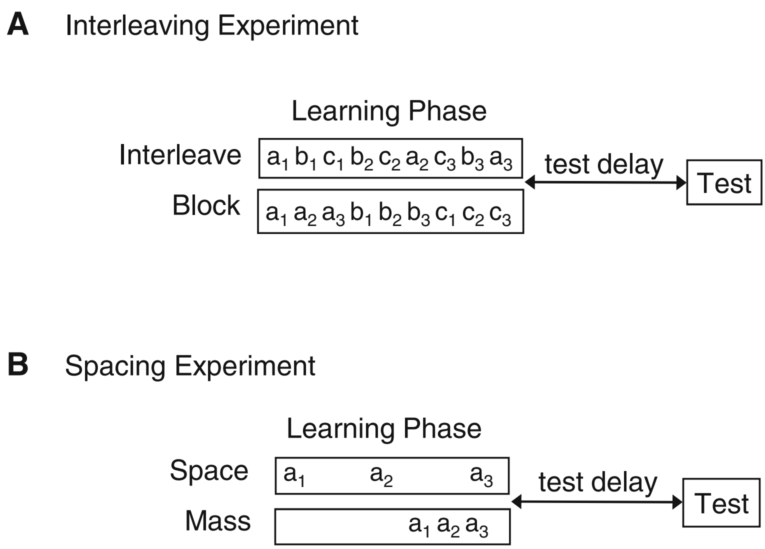

Fig. 1 Prototypical design of interleaving experiment (**a**) and spacing experiment (**b**). In the interleaving study, three exposures to one concept (e.g., three questions on pendulums) are blocked together ($a_1a_2a_3$) or interleaved with exposures to similar concepts (e.g., free fall and incline motion). In the spacing experiment, multiple exposures to one concept are separated by spacing gaps or massed in immediate succession. Neither interleaving nor spacing alters the nature or number of learning exposures, yet both interventions typically improve test scores

change-up, etc.). Interleaving led to superior hitting on a final test requiring batters to hit pitches of each type without knowing the type of pitch in advance, as in a real game.

The remainder of this review is devoted to interleaving. For reasons of clarity, summaries of findings are not contained in a single section but instead distributed through the review. Also, this review provides only an overview of the research on interleaving, and much of the discussion is instead devoted to the potential efficacy and implementation of interleaving in the classroom. A comprehensive review of the interleaving literature can be found elsewhere (e.g., Dunlosky *et al.*, in press).

Category Induction Learning

A recent spate of interleaving research was inspired by a series of studies reported by Kornell and Bjork (2008). In these studies, college students learned to distinguish among the styles of different artists by viewing landscape paintings by each artist. The task was challenging because the artists had similar styles (Fig. 2). In one of the studies (Experiment 1b), for instance, subjects viewed six paintings by each artist, one at a time, and each painting appeared with the artist's name. In the blocked group, paintings were grouped by artist so that subjects first saw all six paintings by Artist A, followed by all six paintings by Artist B, and so forth. The interleaved group saw one painting by A, followed by one painting by B, and so forth, until they had cycled through the list of artists six times. On the final test, subjects saw previously unseen paintings by each artist and tried to select the artist's name from a list. Interleaving improved test scores, 59 vs. 36 %, Cohen's $d=1.28$.

The artist identification task is a category induction task because the paintings by each artist constitute a category, and subjects must learn to identify the category to which a previously unseen painting belongs. This means that subjects cannot perform the task by merely memorizing who painted each of the paintings seen during the learning phase. They must instead learn to recognize, consciously or otherwise, the features of an artist's paintings that distinguish it from paintings by other artists. The features that characterize an artist's style are learned through induction because subjects must generalize from specific instances (i.e., specific paintings).

The benefit of interleaving on category induction has been demonstrated several times in recent years. In one of these studies, for instance, 3-year-old children were shown novel objects one at a time and told each object's name (wug). Similar objects had the same name (e.g., each rattle-like shape was called a "wug"). Children who saw the objects in an order that was interleaved (*wug*, *dax*, *blicket*, etc.) rather than blocked (*wug*, *wug*, *wug*, *dax*, *dax*, *dax*, etc.) were better able to name previously unseen objects on a subsequent test (Vlach *et al.* 2008). More recently, Wahlheim *et al.* (2011) reported two studies in which subjects learned to distinguish among 12 different families of birds (e.g., finches, orioles, and warblers). On each trial, subjects saw two birds belonging to the same family or different families. The simultaneous presentation of two birds from different families is akin to interleaving because members of different categories are juxtaposed, and this different-bird strategy led to superior bird identification on the final test.

Why Does Interleaving Improve Discrimination Learning?

While the results described thus far demonstrate that interleaving can improve students' ability to distinguish among similar concepts, these findings do not point to a particular

Fig. 2 Sample of paintings used in Kornell and Bjork (2008). Subjects saw paintings one at a time. In the blocked study condition, paintings were grouped by artist so that subjects first saw six paintings by Artist A, followed by six paintings by Artist B, and so forth. In the interleaved condition, subjects saw one painting by A, followed by one painting by B, and so forth, until they had cycled through every artist six times. On a final test requiring subjects to identify the artist for previously unseen paintings, interleaving group outscored the blocked group, 59 vs. 36 %

explanation of why this occurs. Two plausible explanations are explored in this section. Although the two accounts are not mutually exclusive, one is not supported by recent data.

By one of the two accounts, the interleaving effect is not an effect of interleaving per se but instead an artifact of spacing, which is a well-known intervention whereby exposures to a single concept are *spaced* apart ($a_1 \ldots a_2 \ldots a_3$) rather than *massed* in immediate succession ($a_1 a_2 a_3$). For example, a reading teacher who ordinarily presents a list of new "sight words" by cycling through the list two times in immediate succession (massing) could instead separate the two cycles one hour apart (spacing). A typical spacing experiment is illustrated in Fig. 1b. The difference between interleaving and spacing is subtle but critical: spacing describes the scheduling of exposures to a *single* concept (A), and interleaving describes the scheduling of exposures to *multiple* concepts (A, B, and C).

Although interleaving and spacing are different interventions, the two are inextricably linked because interleaving inherently introduces spacing. That is, when exposures to multiple concepts are interleaved ($a_1b_1c_1b_2c_2a_2c_3b_3a_3$) rather than blocked ($a_1a_2a_3b_1b_2b_3c_1c_2c_3$), the exposures to any one of the concepts are spaced ($a_1 \ldots a_2 \ldots a_3$) rather than massed ($a_1a_2a_3$). This means that a typical experimental comparison of interleaving and blocking is confounded, and that any benefit of spacing works *in favor* of the interleaving effect. This account is hardly a straw man hypothesis because spacing can dramatically improve learning scores on a final test—a finding known as the *spacing effect*. The spacing effect has been demonstrated in hundreds of studies, including ones with learning materials drawn from the classroom (e.g., Austin 1921; Bahrick and Phelps 1987; Bird 2010; Bloom and Shuell 1981; Carpenter *et al.* 2009; Cepeda *et al.* 2009; Reynolds and Glaser 1964; Metcalfe *et al.* 2007; Seabrook *et al.* 2005).

In order to assess whether the interleaving effect is a spacing effect in masquerade, a few recent studies have compared the effects of interleaving and blocking while controlling for the effect of spacing. In one of these studies, Kang and Pashler (2012) used the artist identification task described above. Subjects saw paintings by three artists (A, B, and C). One group of subjects saw the paintings in an interleaved order ($a_1b_1c_1a_2b_2c_2a_3b_3c_3$ etc.), where b_3 represents the third paintings by Artist B. Another group of subjects saw the paintings in an order that was blocked *and* spaced ($a_1 \ldots . a_2 \ldots . a_3 \ldots . b_1 \ldots . b_2 \ldots . b_3 \ldots . c_1 \ldots . c_2 \ldots c_3 \ldots .$ etc.), with each painting followed by an unrelated cartoon drawing. This allowed the experimenters to ensure that the duration of the spacing gap between any two successive paintings by the *same* artist (such as b_2 and b_3) was the same for all subjects. Despite this control, the interleaving group outscored the blocked-spaced group on the final test (68 vs. 61 %, d=0.78). This same paradigm was used in a mathematics learning study reported by Taylor and Rohrer (2010), which is described in the section on mathematics learning, and they also found an interleaving effect after controlling for the effects of spacing (similar findings have been reported with perceptual learning tasks as well, e.g., Mitchell *et al.* 2008). By disentangling the contributions of spacing and interleaving, these studies demonstrated that interleaving per se, and not the incidental spacing that accompanies interleaving, can sharply improve learning.

If the spacing effect is not responsible for the interleaving effect, how does interleaving improve discrimination learning? The most parsimonious explanation is that interleaving makes it easier for learners to compare and contrast members of one category with members of a different category. Specifically, members of one category (e.g., finches) might differ from members of another category (e.g., sparrows) on a number of dimensions, and the juxtaposition of two members from different categories helps learners appreciate which dimensions (or features) are most relevant to the task of discrimination (e.g., Kang and Pashler 2012; Kornell and Bjork 2008; Mitchell *et al.* 2008; Wahlheim *et al.* 2011).

Mathematics Learning

Interleaving appears to benefit mathematics learning as well, yet most mathematics students devote most of their practice time to blocked practice. This is because each lesson in most mathematics textbooks is followed immediately by a set of practice problems devoted to that lesson. For example, a lesson on ratios might be followed by 20 ratio problems. Not every problem in a blocked assignment is identical in form—for instance, some of the problems might be word problems—but the problems within a blocked assignment are generally based on the same concept or procedure. By contrast, in a few mathematics textbooks, each lesson is followed by an assignment consisting primarily of problems from previous lessons. For example, instead of seeing 20 ratio problems after a lesson on ratios, students might work 5 ratio problems (blocking) and one problem on *each* of 15 earlier topics (interleaving). The displaced ratio problems appear in later assignments so that, by the end of the course, students have seen the same problems as they ordinarily would.

Several mathematics learning studies have compared interleaving and blocking, and the first of these was reported by Mayfield and Chase (2002). In their experiment, college students in need of mathematics remediation attended dozens of sessions over a period of several summer months in which they solved problems using five algebraic rules about exponents (e.g., $ax^{\mathrm{m}} \cdot bx^{n} = abx^{m+n}$). Two of the three subject groups employed practice strategies that essentially provided interleaved and blocked practice (though logistical constraints led to various complicated confounds between these two groups). Subjects were tested 1 or 2 days after the last practice session, and they returned for a second test between 4 and 12 weeks later, depending on their availability. On both tests, the interleaved practice group outscored the blocked practice group by factor of at least 1.3. In short, although procedural complications cloud the interpretation of this study, its results appear to favor interleaving.

Three other studies have directly compared interleaved and blocked mathematics practice. In one of these studies, Rohrer and Taylor (2007) taught college students to find the volumes of four obscure solids (Fig. 3a). Every subject saw the same problems, but the four kinds of practice problems were interleaved or blocked (Fig. 3b). Subjects saw the solution to each practice problem immediately after they tried to solve it. The interleaved group performed *worse* than the blocked practice group during the practice session because the blocked group knew the appropriate strategy in advance. However, interleaving boosted scores on a final test, 63 vs. 20 %, $d=1.34$ (Fig. 3c). In a similar study with the same four kinds of problems, LeBlanc and Simon (2008) also found an effect of interleaving. Finally, in a study reported by Taylor and Rohrer (2010), young students (ages 10 and 11) were taught to solve four kinds of mathematics problems relating to prisms (e.g., if the base of a prism has 9 sides, how many edges does the prism have?), and a test 1 day later revealed an interleaving effect, 77 % vs. 38 %, $d=1.21$. Further analysis revealed that this test benefit of interleaving was due *solely* to the fact that blocking group made more errors of the kind in which a student used the strategy that was appropriate for one of the *other* three kinds of problems. In other words, the entire interleaving effect was due to the elimination of discrimination errors.

These findings suggest that interleaved mathematics practice helps students learn to distinguish between different kinds of problems. This is a critical skill because solving a mathematics problem requires that students first identify what kind of problem it is, which means that they must identify those features of a problem that indicate which concept or procedure is appropriate (e.g., Crowley *et al.* 1997; Kester *et al.* 2004; Siegler and Shrager 1984). Identifying the kind of problem is not always easy. For example, a large portion of algebra is devoted to solving equations, but the instruction, "Solve for *x*," does not indicate which one of several solving strategies is appropriate. For instance, students must use the quadratic formula to solve some equations (e.g., $x^2-x-1=0$), and they must factor to solve others (e.g., $x^3-x=0$). As noted in the first section of this paper, this kind of discrimination task is essentially a categorization task because students must identify the category of problems to which a particular problem belongs (e.g., an equation that can be solved by factoring). In other terms, solving a mathematics problem requires students to know *which* strategy is appropriate and not only *how* to execute the strategy.

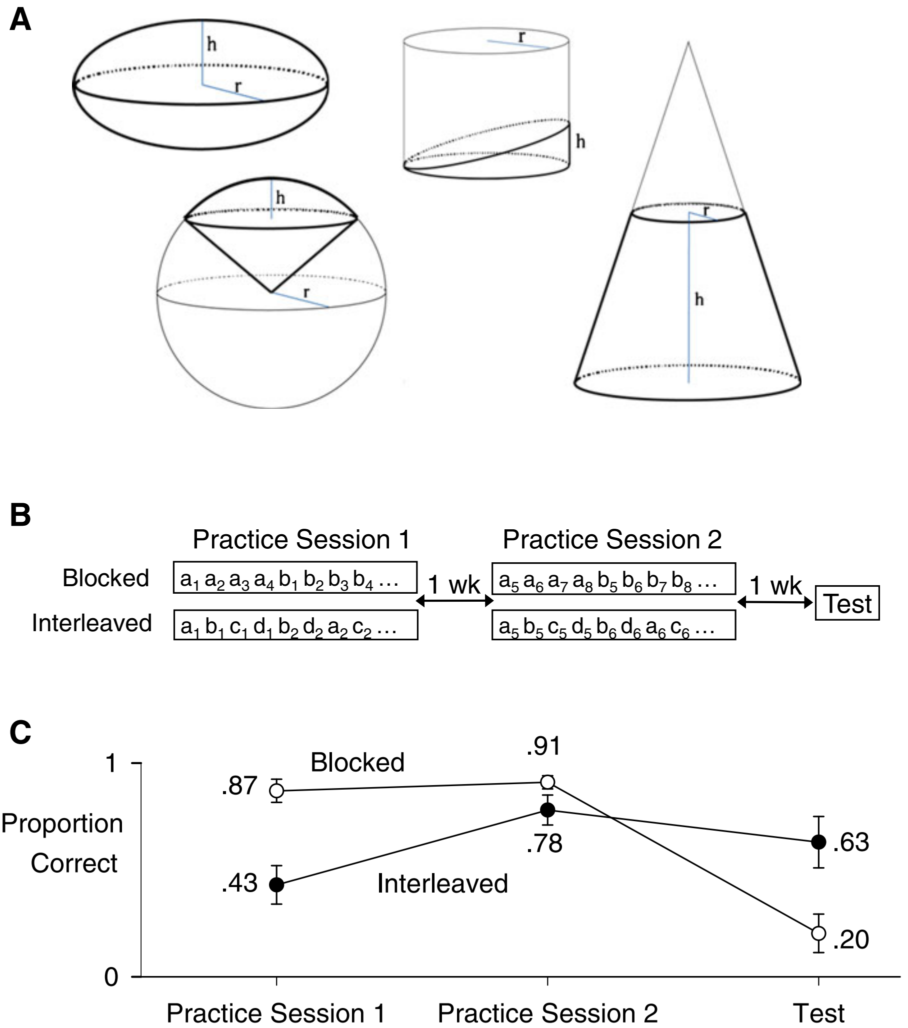

Fig. 3 The stimuli (**a**), procedure (**b**), and results (**c**) of a mathematics learning experiment (Rohrer and Taylor 2007). Subjects solved four different kinds of problems (*a*, *b*, *c*, and *d*), and only the order of the problems was varied. Interleaving reduced practice scores yet tripled test scores (d=1.34). *Error bars* indicate plus or minus one standard error

With blocked practice, however, students need not identify an appropriate strategy because every problem in the assignment can be solved by the same strategy. For example, if a statistics assignment includes a dozen problems requiring students to assess the statistical significance of data obtained with a particular kind of research design, and every problem requires the same statistical test (e.g., repeated-measures *t* test), students know the appropriate statistical test in advance. This is problematic because students should ultimately learn how to select the appropriate statistical test on the basis of the research design, and this skill is arguably more important than knowing how to perform the statistical test. In essence, blocking provides scaffolding. This might be useful when students see a new kind of problem, but students who receive only blocked assignments do not have the opportunity to practice without this crutch. Of course, when students see problems during a cumulative exam, standardized test, or subsequent course, the crutch is snatched away.

In essence, the pedagogical value of a practice problem is altered when the problem appears within a blocked assignment, as illustrated by the following word problem from a popular textbook (p. 99; Carter *et al.* 2011):

> Rhode Island is the smallest state in the United States. Its area is about 1/6 the area of New Hampshire. If the area of New Hampshire is about 9,270 square miles, what is the approximate area of Rhode Island? Answer $= (1/6) \times 9270 = 1545$ square miles

This problem is solved by multiplying the two given numbers, and almost all middle school students can multiply with the aid of a calculator (which is allowed during most standardized tests). However, this problem is challenging because students must first infer that they *should* multiply. Yet no such inference is needed if the problem appears immediately after a group of problems explicitly requiring the strategy (i.e., multiply the two given numbers), which is exactly where this problem appears (Fig. 4). In other words, blocked practice allows students to solve this *word* problem without reading any words and yet most word problems appear immediately after a group of problems requiring the same procedure. This might partly explain the notorious difficulty of word problems (Rohrer 2009).

In summary, proficiency in mathematics and certain sciences requires that students learn to choose an appropriate strategy for each kind of problem. Students have an opportunity to practice this skill when problems of different kinds are interleaved, whereas blocked practice allows students to safely assume that each problem can be solved with the strategy used to solve the previous problem. A number of mathematics learning studies have shown that interleaving (rather than blocking) improves scores on final tests of learning, yet the vast majority of problems in most mathematics textbooks appear within a blocked assignment.

Glencoe McGraw-Hill (Grade 8)

Multiply. Write in simplest form.

1. $(^3/_5) \cdot (^5/_7)$	2. $(^4/_5) \cdot (^3/_8)$	3. $(^6/_7) \cdot (^7/_6)$
4. $(-\,^1/_8) \cdot (^4/_9)$	5. $(-\,^2/_9) \cdot (^3/_8)$	6. $(-\,^{12}/_{13}) \cdot (-\,^2/_3)$
7. $(1\,^1/_3) \cdot (5\,^1/_2)$	8. $(2\,^1/_2) \cdot (1\,^2/_5)$	9. $(-\,6\,^3/_4) \cdot (1\,^7/_9)$

10. Rhode Island is the smallest state in the United States. Its area is about 1/6 the area of New Hampshire. If the area of New Hampshire is about 9,270 square miles, what is the approximate area of Rhode Island?

Fig. 4 Excerpt of blocked assignment from Grade 8 mathematics textbook. After students solve problems 1–9, which explicitly require the multiplication of the two given numbers, they can safely assume that Problem 10 also demands that they multiply the two given numbers. This illustrates that blocked practice sometimes allows students to solve a word problem without reading any words. If Problem 10 had instead appeared within an interleaved assignment, surrounded by different kinds of problems, students would need to infer the appropriate strategy on the basis of the problem itself (excerpt from Carter *et al.* 2011, p. 99)

Caveats and Limitations

Although the studies reviewed here uniformly support interleaving, there are several reasons to be cautious about its utility in the classroom. Most notably, studies of interleaving have not employed ecologically valid procedures. For instance, in most of the studies cited here, the learning phase was limited to a single session, and the delay between the learning phase and the test was less than 1 h. The case for interleaving requires evidence from classroom-based experiments with educationally meaningful procedures because many promising interventions have fizzled in the classroom.

Another shortcoming of the literature is the absence of studies assessing a combination of blocking and interleaving, which might be the optimal strategy for complex learning materials. For example, the first portion of each mathematics assignment might include a small block of problems on the concept or procedure learned that day, with the remainder of the assignment devoted to problems based on previous lessons. This hybrid approach might find greater acceptance in the classroom because students and teachers are accustomed to blocking.

Another caveat is that teachers and textbooks already employ interleaving for some kinds of material. For example, some sets of easily confused terms (e.g., solute, solvent) are inherently juxtaposed in instructional materials because the terms are intrinsically related (a solute is a substance dissolved in a solvent). Likewise, some pairs of similarly spelled words (allusion–illusion or assent–ascent) are commonly presented to students as a pair so that the distinction is more salient.

The most notable limitation of interleaving, however, is that its benefits are likely limited to difficult discriminations. The learning tasks in the studies cited in this review were obviously chosen to produce discrimination errors (e.g., the artist paintings in Fig. 2), and discriminations of this subtlety are rarely encountered in the classroom. If these studies had instead required subjects to distinguish between *dissimilar* concepts or terms, the size of the interleaving effects would have almost certainly been smaller. This boundary condition sharply constraints the utility of interleaving.

Feasibility of Implementation

In addition to questions surrounding the generality and efficacy of interleaving, little is known about its feasibility. The successful implementation of any intervention depends on factors other than its efficacy, including its cost, ease of use, and perceived efficacy. Each of these factors is considered here.

Interleaving is not expensive. Interleaving can be adopted without changes to curricula, and creators of learning materials can incorporate interleaving by rearranging examples, questions, or problems. For instance, in revising a textbook for its next edition, the lessons can remain intact, and one or two questions could be drawn from each blocked assignment and used for interleaved assignments. Consequently, the cost and effort of creating interleaved assignments can be shouldered by publishers rather than teachers and taxpayers.

On the other hand, students might balk because interleaving increases the difficulty of a question or problem. A group of questions or problems are easier when all relate to the same topic or concept. By contrast, answering a set of interleaved biology questions might require students to consult material presented in previous chapters, and an interleaved mathematics assignment prevents students from simply repeating the same procedure throughout the assignment. Students will therefore make more errors, and work more slowly, when assignments are interleaved. This in itself is not problematic because the aim of classroom instruction is ultimate mastery, not error-free learning. Still, some students might be unwilling to make the extra effort. In this scenario, interleaving is like bad-tasting cough syrup—ineffective because children refuse to use it. Future research might examine how the likability of interleaving can be improved. For instance, each question in an interleaved biology assignment could include the page number of a relevant example or definition appearing earlier in the text.

The added challenge imposed by an interleaved order might also lead students and teachers to falsely yet reasonably conclude that interleaving is less effective than blocking. Although the relative ease of blocked assignments is due to the fact that blocking provides scaffolding, students and teachers may not be aware of this. Consequently, blocking leads students to believe that they understand material better than they actually do—what Son and Kornell (2010) have called an illusion of knowing. These kinds of poor *metacognitive* judgments have been demonstrated many times (e.g., Dunlosky and Lipko 2007; McCabe 2011; Metcalfe 2000).

In fact, several studies have shown that learners doubt the efficacy of interleaving even after they have tried it. For instance, in one of the artist learning studies (Kornell and Bjork 2008), subjects relied on blocking for some artists and interleaving for others, and, immediately after the final test, they were asked to indicate which one of the two learning strategies, if either, "helped them learn more." Among subjects who did benefit from interleaving, only 25 % believed that interleaving was more helpful. Interleaving also failed to impress subjects in the earlier-described bird learning study (Wahlheim *et al.* 2011), although the details of their analysis are beyond the scope of this review. Finally, in a mathematics learning study reported by LeBlanc and Simon (2008), interleaving roughly doubled test scores, yet subjects' predictions of their test scores, made immediately *after* their last practice problem, were barely affected by whether their practice was interleaved or blocked.

The broader point is that the feasibility of an intervention depends partly on whether students, teachers, and creators of learning materials are convinced of its benefits. This is more problematic than it should be because instructional methods are often chosen with little regard for evidence. For instance, more than a century of research has shown that spacing is an extremely effective learning strategy, yet massing remains more popular in the classroom despite numerous appeals (e.g., Bjork 1979; Cepeda *et al.* 2008; Dempster 1989; Halpern 2008; Rohrer and Pashler 2007; 2010; Schwartz *et al.* 2011; Willingham 2002). As many others have argued, teachers, those who train teachers, and researchers need a greater appreciation for experimentally supported instructional methods (e.g., Robinson *et al.* 2007; Slavin 2002).

Summary

When students must learn to distinguish among similar concepts (or terms or principles or kinds of problems), the findings reviewed here suggest that the exposures to each of the concepts should be interleaved rather than blocked (Fig. 1a). For certain kinds of materials (such as the paintings in Fig. 2b), this interleaving effect may occur because interleaving guarantees that exposures to different concepts are juxtaposed, making it easier for learners to identify the feature that distinguishes members of one category from another. With mathematics problems, interleaved practice requires students to choose the appropriate strategy for a problem because each problem is different from the previous one.

In spite of these results, there are reasons to be cautious about the utility of interleaving in the classroom. The findings reviewed here were obtained in laboratory settings with procedures lacking ecological validity, and long-term, classroom-based studies are needed before interleaved can be recommended without qualification. There are also questions about its feasibility. Although the financial costs of implementing an intervention are relatively inexpensive because learning materials need only be rearranged, interleaved assignments can be more challenging for students because they do not know the relevant concept or procedure before reading each question or problem.

In brief, the limited evidence prohibits a wholesale endorsement of interleaving. At the same time, however, it seems foolhardy for students to rely solely on blocking when the data clearly favor interleaving over blocking. As a first step, then, it seems prudent to recommend that at least a portion of the exposures to related concepts be interleaved.

Acknowledgments I thank Pooja Agarwal and one anonymous reviewer for their comments on an earlier version of this review. This work was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110517 to the University of South Florida (PI: D. Rohrer). The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

References

Austin, S. D. M. (1921). A study in logical memory. *The American Journal of Psychology, 32*, 370–403.

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

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied PsychoLinguistics, 31*, 635–650.

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

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

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U. S. history facts. *Applied Cognitive Psychology, 23*, 760–771.

Carter, J. A., Cuevas, G. J., Day, R., Malloy, C. E., Kersaint, G., Luchin, B. M., et al. (2011). *Math Connects Plus: Course 3 (Florida Version). School Education Group (SEG), McGraw-Hill, Glencoe*. New York: McGraw-Hill.

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.

Cepeda, N. J., Mozer, M. C., Coburn, N., Rohrer, D., Wixted, J. T., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56*, 236–246.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Crowley, K., Shrager, J., & Siegler, R. S. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. *Developmental Review, 17*, 462–489.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*, 309–330.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228–232.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (in press). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*.

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

Halpern, D. F. (2008, March). *25 learning principles to guide pedagogy and the design of learning environments*. Paper distributed at the keynote address at the Bowling Green State University Teaching and Learning Fair. Bowling Green, OH.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*, 97–103.

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2004). Timing of information presentation in learning statistics. *Instructional Science, 32*, 233–252.

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

LeBlanc, 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.

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

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

Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 197–211). London: Oxford University Press.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology, 19*, 743–768.

Mitchell, C. J., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 237–242.

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.

Robinson, D. H., Levin, J. R., Thomas, G. D., Pituch, K. A., & Vaughn, S. R. (2007). The incidence of "causal" statements in teaching and learning research journals. *American Educational Research Journal, 44*, 400–413.

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

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*, 183–186.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406–412.

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

Schwartz, B. L., Son, L. K., Kornell, N., & Finn, B. (2011). Four principles of memory improvement: A guide to improving learning efficiency. *International Journal of Creativity and Problem Solving, 21*, 7–15.

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

Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. In J. Kilpatrick, G. W. Martin, & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 119–233). Reston: National Council of Teachers of Mathematics.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229–293). Hillsdale: Erlbaum.

Skinner, B. F. (1933). The rate of establishment of a discrimination. *The Journal of General Psychology, 9*, 302–350.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.

Son, L. K., & Kornell, N. (2010). The virtues of ignorance. *Behavioral Processes, 83*, 207–212.

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

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*, 163–167.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*, 750–763.

Willingham, D. T. (2002). Allocating student study time: Massed vs. distributed practice. *American Educator, 47*, 37–39. Summer.