Journal of Experimental Psychology: Applied

ISSN: 1076-898X

2022, Vol. 28, No. 1, 100–113
https://doi.org/10.1037/xap0000391

Scheduling Math Practice: Students' Underappreciation of Spacing and Interleaving

Marissa K. Hartwig[1], Doug Rohrer[1], and Robert F. Dedrick[2]
[1] Department of Psychology, University of South Florida
[2] Department of Educational and Psychological Studies, University of South Florida

Many randomized controlled experiments in the classroom have found that mathematics learning is improved dramatically when practice problems of one kind are distributed across multiple assignments (spaced) and mixed with other kinds of problems (interleaved). In two studies, we investigated students' knowledge of spacing and interleaving. In Study 1, 193 undergraduates designed learning schedules for a hypothetical math class. In Study 2, 175 undergraduates selected from among five hypothetical schedules in response to a variety of questions, provided reasons for their selections, and rated the utility of spacing and interleaving. In both studies, most participants incorrectly judged schedules with minimal degrees of spacing and interleaving to be most effective. Also, schedules with more spacing and interleaving were perceived as more difficult, less enjoyable, and less common. Participants' ratings of utility revealed mixed perspectives on spacing and an underappreciation of interleaving. Altogether, these findings demonstrate that most students fail to recognize the benefits of spaced and interleaved practice. Further, by identifying specific ways in which their beliefs about spacing and interleaving fall short, we reveal opportunities to reshape students' beliefs to foster these effective learning techniques.

Public Significance Statement

Many studies have shown that spacing (i.e., distributed practice of the same concept across time) and interleaving (i.e., practicing a mix of concepts within a study session) are highly effective learning techniques. The studies reported here examined college students' beliefs about these techniques in the context of math learning—a familiar domain in which these techniques are highly effective and easy to implement. Even in this context, we found that most college students did not grasp the benefits of spacing or interleaving. We also identified specific errors and gaps in their beliefs, thereby revealing opportunities to correct their beliefs and encourage them to utilize these highly effective learning techniques.

Keywords: scheduling, math, interleaving, spacing, beliefs

An abundance of research has identified learning techniques that reliably boost student learning (for reviews, see Carpenter, 2014; Dunlosky et al., 2013; Kang, 2016; Roediger & Pyc, 2012). Among these effective techniques are *spacing* (i.e., distributed practice of the same concept across time) and *interleaving* (i.e., practicing a mix of concepts interspersed within a study session). In mathematics learning, the focus of this paper, several randomized controlled trials have found large benefits of spaced practice (a single kind of practice problem distributed across assignments) and interleaved practice (different kinds of math problems mixed within the same assignment; e.g., Hopkins et al., 2016; Lyle et al., 2020; Rau et al., 2013; Rohrer, Dedrick, Hartwig, et al., 2020). Yet no one has investigated whether students recognize the utility of these techniques for math learning. For example, do students falsely believe that practice problems of the same kind should be mostly concentrated in one practice session? Students' beliefs about the effectiveness of learning techniques like spacing and interleaving can influence their study decisions and thereby profoundly affect learning outcomes, especially when students must manage their own study (see theories of self-regulated learning, e.g., Winne & Hadwin, 1998). Indeed, college students must make many choices about when they study and how they practice, so their beliefs about learning techniques are consequential. Thus, the present research investigated college students' beliefs about spacing and interleaving—that is, their *metacognitive knowledge* about these learning techniques.

The importance of metacognitive knowledge is recognized by various theories of metacognition and self-regulated learning. A classic framework of metacognition by Flavell (1979) proposed four categories of metacognitive phenomena: metacognitive knowledge (e.g., knowledge or beliefs about learning strategies, task demands, or the self as a learner), metacognitive experience (e.g., in-the-moment perceptions of learning progress), goals (or tasks), and actions (e.g., study decisions). Theories of self-regulated learning

This article was published Online First January 6, 2022.
Marissa K. Hartwig https://orcid.org/0000-0003-3692-9105
Robert F. Dedrick https://orcid.org/0000-0001-5696-750X
This research was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A160263 to the University of South Florida. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.
Correspondence concerning this article should be addressed to Marissa K. Hartwig, Department of Psychology, University of South Florida, 4202 East Fowler Avenue, Tampa, FL 33620, United States. Email: mhartwig@usf.edu

are broader in scope, positing a mix of cognitive, metacognitive, behavioral, motivational, social, emotional, and contextual factors that affect learning (for a recent review, see Panadero, 2017). Importantly, virtually all these theories recognize that metacognitive knowledge can affect study behavior and, in turn, learning (e.g., Borkowski et al., 2000; Pintrich, 2000; Winne & Hadwin, 1998). For instance, a student's metacognitive knowledge about a learning technique, such as interleaving, can affect whether the student is willing or able to effectively use the technique. These theories suggest that metacognitive knowledge can be acquired through experience as well as direct instruction; and metacognitive knowledge can be accurate or inaccurate, detailed or sparse, domain-specific or general. Further, metacognitive knowledge about learning techniques can include declarative knowledge (about different techniques and their effectiveness), procedural knowledge (about how to implement a technique), and conditional knowledge (about when or why a technique is appropriate; e.g., Pintrich et al., 2000). These theories offer useful frameworks for understanding metacognitive knowledge, but they do not predict whether students possess knowledge of spacing or interleaving, per se, or whether their beliefs are accurate.

Examining students' knowledge of specific learning techniques—especially techniques demonstrated to be efficacious (such as spacing and interleaving)—can help determine whether students possess these techniques in their "toolbelt" of strategies or whether their metacognitive knowledge needs improvement. Several surveys suggest that college students often do not know or use good study techniques (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007). Some recent studies have explored whether students recognize the utility of specific techniques like spacing or interleaving in certain contexts (e.g., Susser & McCabe, 2013; Yan et al., 2016), and students' appreciation of these techniques seems to be limited (further detail below). But would students be able to recognize the utility of spacing and interleaving in the context of math learning? Math courses typically feature many practice problems—that is, small, distinct units of practice that can easily be spaced or interleaved (compared to lectures or more lengthy activities that may be awkward to split into small units). Furthermore, college students have considerable familiarity with math learning across years of schooling, and their experience with math learning could contribute to metacognitive knowledge in this domain. Thus, exploring students' beliefs about math learning may allow a more realistic assessment of whether college students possess any awareness of the benefits of spacing and interleaving.

Importantly, students' beliefs about spacing and interleaving may be nuanced, especially with respect to implementation. For instance, implementation of spacing is more than a yes–no decision to either space or not space; rather, students must also decide how and when to space their study. These implementation details can reveal the nuances of students' beliefs and how those beliefs might be improved. Thus, the present studies measured students' beliefs with a task we believe to be more informative than a typical survey. Surveys often require participants to make generalizations about their studying (i.e., to aggregate their behavior or beliefs across time and across differing circumstances, activities, and subject areas), which can make participants' responses hard to interpret and possibly less accurate. In contrast, in the present studies, we asked participants to design or select schedules of math practice for a given scenario. An analysis of the schedules revealed college students' intuitive beliefs about spacing and interleaving and, importantly, allowed us to quantify the amount of spacing and interleaving the students believed to be most effective. Before we describe the present studies, we first clarify the distinction between spacing and interleaving, briefly summarize the empirical literature for both techniques, and review research relevant to students' beliefs about the techniques.

Spacing Versus Interleaving

Each concept in a math course is seen by students on more than one occasion, and these encounters can be distributed across only a few days (less spacing) or spread more thinly across many days (more spacing). For example, the concept of slope could be taught in Monday's lesson and revisited with practice problems on the following day and perhaps again during a review before an exam, or the same number of practice problems could be distributed across dozens of days. As another example, 12 slope problems might be concentrated in only two assignments (six problems each) or distributed across four assignments (three problems each). Importantly, the notion of spacing pertains to the scheduling of one topic or concept, such as slope, across time.

Apart from how exposures to a single topic are scheduled, teachers and textbook designers must decide how to arrange problems relating to different topics. Most commonly, students see a block of problems devoted to the same concept or skill. Alternatively, an interleaved practice assignment might include a mix of problems (e.g., one problem on slope, followed by one problem on area, and so forth) such that students do not know in advance which concept or skill will be required by the next problem. Whenever different kinds of practice problems are interleaved, any one kind of problem is necessarily spaced. Nonetheless, spacing and interleaving are distinct. Indeed, practice can be spaced without interleaving problem types (e.g., a block of slope problems each Friday in March, a block of area problems each Friday in April, and so forth).

Spacing and Interleaving Boost Learning

Spacing usually improves scores on delayed tests (e.g., on a test given a week after the last practice session), and this benefit is one of the most thoroughly studied and well-established effects in learning research (for reviews, see Carpenter, 2014; Cepeda et al., 2006; Delaney et al., 2010; Dempster, 1989; Dunlosky et al., 2013; Kang, 2016). Countless studies have demonstrated that spaced practice produces better performance on delayed tests compared to practice that is less spaced or not spaced at all (massed), and the effect sizes are often large. Importantly, spacing effects occur even though the total time spent practicing remains the same and the test delay (i.e., the delay between the last practice session and the test) is held constant. Thus, spacing boosts test performance due to the schedule itself, not because students spent more time studying or studied more recently. Researchers have proposed numerous theoretical explanations for the spacing effect (for reviews, see Benjamin & Tullis, 2010; Delaney et al., 2010; Dempster, 1989). According to various theories, the spacing effect may derive from mechanisms such as encoding variability (i.e., contextual variation provides richer encoding when two learning episodes are spaced apart), deficient processing (i.e., processing of material during a second learning episode is diminished if close in time to the first episode), consolidation (i.e., a second learning episode benefits from any memory consolidation that occurs in the interim), or study-phase retrieval (i.e., spacing promotes effortful retrieval during a second learning episode). However, no single mechanism has accounted for the entire body of spacing-related findings, and it is possible that a combination of mechanisms may best explain the effect (Delaney et al., 2010).

Regardless of mechanism, spacing effects are robust—occurring across various materials, procedures, and learner characteristics (Dunlosky et al., 2013). Most important for the present study, spacing effects have been demonstrated in numerous classroom-based randomized studies (e.g., Seabrook et al., 2005; Sobel et al., 2011; for a review, see Dunlosky et al., 2013). Moreover, classroom studies have found spacing effects with math learning (Barzagar Nazari & Ebersbach, 2019; Hopkins et al., 2016; Lyle et al., 2020; Schutte et al., 2015). In short, considerable data show that spaced math practice improves scores on delayed tests. Less is known about how to optimize schedules of spaced practice, which is a topic of continued investigation (e.g., Storm et al., 2010) and likely depends on many factors (e.g., number of sessions, time between sessions, total amount of practice, test delay). Nonetheless, the literature is clear that practice should be spaced across many class sessions if students are to retain the information long term (Rawson et al., 2013, 2018).

The benefit of interleaved practice has also received attention in recent years. Research shows that a greater degree of interleaved practice (vs. mostly blocked practice) often produces better scores on subsequent tests. Interleaving has been shown to benefit category induction learning (e.g., Kornell & Bjork, 2008; Vlach et al., 2008), science learning (e.g., Eglington & Kang, 2017), foreign language learning (e.g., Pan et al., 2019), and complex decision-making (e.g., Helsdingen et al., 2011a, 2011b). Also, most relevant here, interleaved practice improved math learning in each of several studies in both the laboratory and classroom (Foster et al., 2019; Mayfield & Chase, 2002; Rau et al., 2013; Rohrer, Dedrick, Hartwig, et al., 2020; Sana et al., 2017). Importantly, a greater degree of interleaved practice improves test scores even though participants in interleaved and blocked groups receive the same practice problems, so the only difference is the order in which problems occur. The benefit of interleaved math practice may be partly attributable to its inherent spacing of concepts and retrieval practice during problem solving (see Rohrer, Dedrick, Hartwig, et al., 2020, for discussion of these contributions), yet evidence suggests interleaving is effective in its own right (Kang & Pashler, 2012; Taylor & Rohrer, 2010). Interleaved practice juxtaposes problems that target different concepts, and this temporal juxtaposition is believed to encourage students to notice the similarities and differences between problem categories and enhance their ability to discriminate those categories (e.g., Rohrer, 2012; but see Foster et al., 2019). Also, during interleaved practice, students benefit from not knowing in advance what strategy will be required by the next problem. That is, blocked practice allows students to simply use the most recent strategy over again, whereas interleaved practice requires students to recognize when a problem calls for that strategy—as they must do when taking cumulative exams or standardized tests.

Students' Beliefs About Spaced and Interleaved Practice

A variety of studies have begun to reveal college students' beliefs about spaced and interleaved practice. Study behavior (either self-reported or observed) may signal whether participants believe these techniques have utility for learning. For example, when participants in laboratory studies were given the choice to space or mass their practice of to-be-learned word pairs prior to a final test (e.g., Pyc & Dunlosky, 2010; Toppino & Cohen, 2010), participants chose spacing over massing under many conditions. Yet, when test delay was controlled and participants were asked to select between longer and shorter spacing, shorter spacing was preferred (Cohen et al., 2013). Also, when researchers tracked when and how often college students chose to study across a semester, students tended to cram before deadlines rather than space (e.g., Taraban et al., 1999). Importantly, though study behaviors may partly signal students' underlying beliefs about effective studying, behaviors are also shaped by practical constraints. For example, even a student who is aware of the benefits of spacing or interleaving may not use them when faced with time pressure or lagging motivation. Furthermore, the presence of behaviors like spacing can occur for reasons that do not signify an understanding of the benefits. For example, students might space their practice if they become bored while working problems in a single session and decide to continue later. In short, students' study choices can be complex, and students likely consider many factors (e.g., amount of material, test difficulty) when deciding how to distribute their study (Susser & McCabe, 2013).

Another way to investigate students' beliefs about spacing and interleaving is to directly survey those beliefs rather than measure behaviors. In one survey, 85% of college students indicated that spaced study (rather than massed) is better for long-term retention of materials (Susser & McCabe, 2013). In another survey, 81% of college students said that flashcards should be spaced rather than massed (Wissman et al., 2012). But what do students envision when they endorse spacing? To find out, researchers have asked participants to describe how or when students should study to effectively learn course material (e.g., Blasiman et al., 2017; Susser & McCabe, 2013; Taraban et al., 1999) and have shown that although many participants endorse some degree of spacing, they nevertheless describe effective study to be heavily concentrated near exams.

Compared to spacing, the utility of interleaving may be harder for students to recognize. Several laboratory studies have asked participants to evaluate interleaving after using both interleaved and blocked practice (Kornell & Bjork, 2008; Kornell et al., 2010; Yan et al., 2016). In these studies, college students practiced classifying paintings by artist (a category induction task), completed a test with novel stimuli, and finally reported which schedule had been more effective. Many participants said that blocked practice was more beneficial than interleaved practice, even when their own test scores showed the opposite effect. Whether students might recognize the utility of interleaving for more common academic tasks, like math practice, remains an open question.

Yet another approach to understanding students' beliefs involves presenting hypothetical scenarios or vignettes. In one study, when given a vignette about spacing, 69% of students correctly rated spaced practice as more effective than massed practice (Morehead et al., 2016). In contrast, when the same students read a vignette about interleaving (using the artist painting paradigm described above), only 16% correctly rated interleaved practice as more effective than blocked practice. With the same interleaving vignette, less than 10% of students in another sample endorsed interleaving over blocking (McCabe, 2011), and even university students training to become teachers underrated the efficacy of interleaving (Halamish, 2018). Math vignettes have not been used, however.

Finally, in an extension of the vignette approach, participants can be asked to create a study schedule for a scenario. Schedules may provide unique insight because they represent a concrete instantiation of the beliefs students hold. In one study by Wissman et al. (2012), participants were shown a calendar for the month of February and asked to imagine that the date was February 1st and that they would take an exam (in General Psychology) at the end of the month. Participants indicated which days they would study, and what those study activities would be, to earn an A on the exam. Similarly, Cohen et al. (2013; Experiment 7) asked participants to imagine they had 1 week to prepare for an exam, had only 12 hr of study time available, and needed to schedule those 12 hr to maximize their exam grade. Both studies found that participants underutilized spacing and instead scheduled much of their study just before the test. Similar scheduling tasks have been used with interleaving. Yan et al. (2017) asked participants to imagine trying to learn the styles of several artists to prepare for an exam in which they must identify novel paintings by those artists. The participants selected the sequence in which they would study paintings during a hypothetical practice session. Blocked sequences were often preferred, whereas heavily interleaved sequences were unpopular. Yan and Sana (2021) asked participants how they would schedule their study of unrelated domains (e.g., geography, math, psychology, history) and concepts within a domain (e.g., integration, volume, geometry, factorial equations) and found that participants did not choose optimal sequencing. Although math was one of several domains in their study, they did not focus specifically on math, nor did their scenario involve the scheduling of practice problems, which are a central activity in math learning.

Though the studies above suggest that learners have a poor appreciation of spacing and interleaving, we wanted to give students a fair chance to display their beliefs for an authentic educational scenario, focusing on a domain that is familiar to them and that typically features many practice problems. Practice problems are prevalent in math (and also in other fields such as science or engineering) and offer a realistic context in which spacing and interleaving might occur. That is, practice problems provide small, distinct units of practice that could easily be interleaved within a session or spaced across many sessions. Thus, students might find spacing and interleaving to be more feasible with practice problems than with other kinds of activities (such as lessons or lectures) that can be awkward to divide into small units. Furthermore, most students study math every year of their schooling from early childhood until college, and in those math classes, virtually all students have experienced some degree of spaced practice (e.g., revisiting math concepts across class sessions or grade levels) and interleaving (e.g., mixed review before exams). This experience might contribute to their metacognitive knowledge for math learning. Math learning is both conceptual and computational, seemingly different from paired-associate learning, fact memorization, or category induction, and beliefs about spacing and interleaving might depend on the type of learning or domain. Thus, the present studies used a math scheduling scenario that included multiple topics and numerous practice problems so we could assess learners' beliefs about both spacing and interleaving of math practice.

The Present Studies

In two studies, we investigated college students' metacognitive knowledge of spacing and interleaving in the context of math learning. In Study 1, participants scheduled lessons and practice problems for a hypothetical math class, and we measured the extent to which a single concept was spaced across class sessions and the extent to which problems of different kinds were interleaved within class sessions. In Study 2, a new sample of participants selected one of several hypothetical schedules in response to a variety of questions designed to gauge their beliefs. In short, we examined whether students' scheduling decisions demonstrate that they appreciate the utility of spacing and interleaving for math learning, and we also examined the beliefs and rationale underlying their scheduling choices.

Study 1

Method

Participants

We tested 193 undergraduate students (121 women, 71 men, 1 unreported) from the psychology participant pool at the University of South Florida. The sample size was large enough to give margins of error smaller than 4% when measuring spacing or interleaving, at 95% confidence. Most students in the participant pool were enrolled in introductory psychology, and they had various majors. Students received course credit for research participation. The sample included 37% freshmen, 23% sophomores, 25% juniors, 14% seniors, and 1% other (e.g., nondegree seeking); the mean age was 20.3 years ($SD = 4.1$, range 18–62); and they self-identified as 65% White, 18% Black, 18% Asian, 3% other (8% unreported), and also 25% Hispanic (2% unreported).

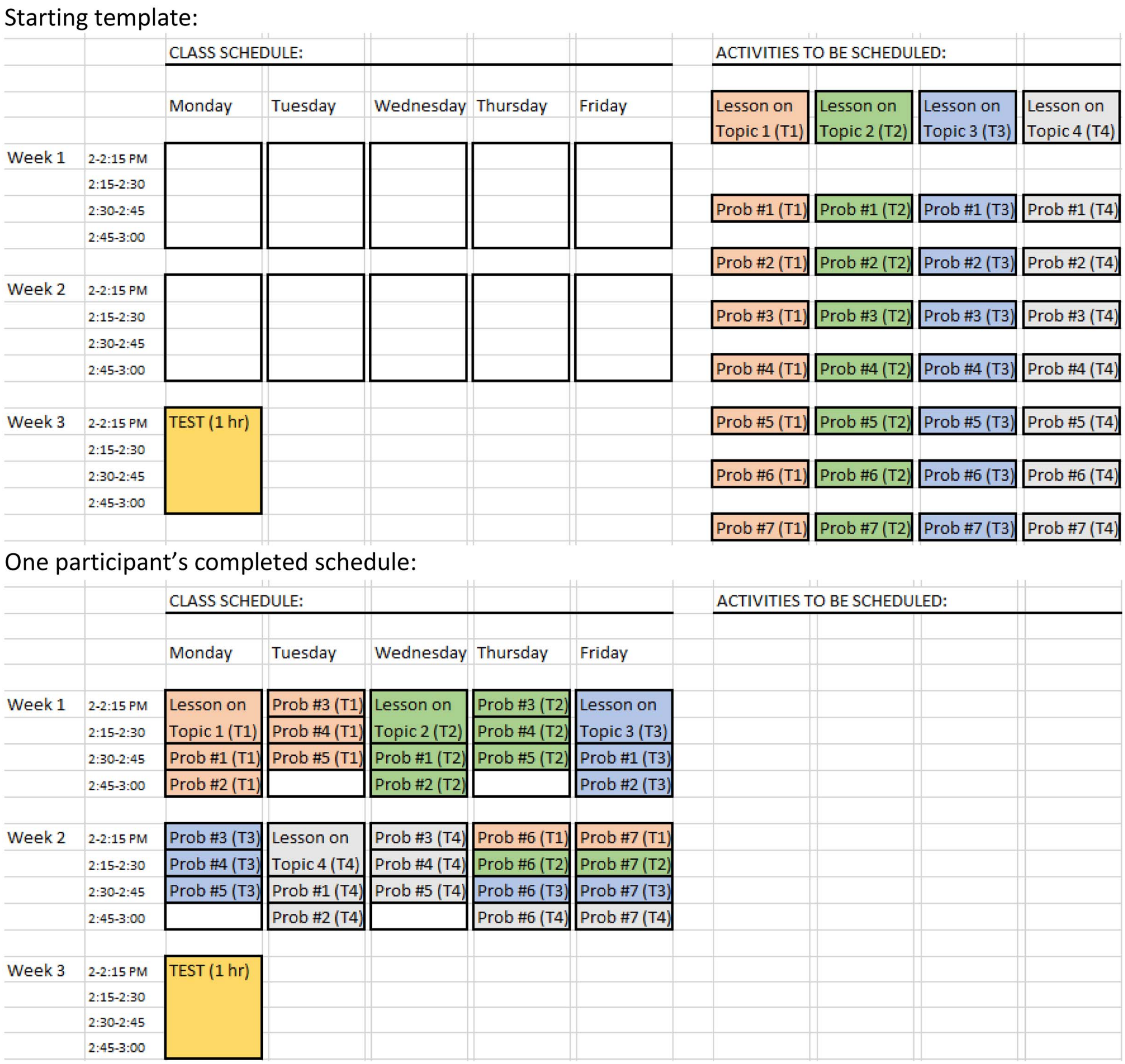
Procedure

Each participant was tested alone in a small room. A computer presented detailed instructions, a blank 2-week schedule for a hypothetical math class, and colored boxes representing four lessons (one per topic) and 28 practice problems (seven per topic) that needed to be scheduled (Figure 1). Topic numbers (1–4) and problem numbers (1–7) were arbitrary labels. The experimenter orally explained the instructions shown on the screen and, when needed, taught participants how to use keyboard shortcuts for cutting and pasting. The instructions asked participants to imagine they were a math teacher designing a class schedule for 2 weeks with the goal of maximizing student performance on a test occurring in week 3. Participants were asked to imagine that each lesson would require 30 min and each problem would require 15 min. Participants were told that no schedule would be considered right or wrong and that the goal of the research was to understand participants' perspectives on learning. Participants worked at their own pace until they had completed the schedule to their satisfaction. Most participants completed the task in about 10 min.

Results and Discussion

Participants created a variety of schedules. As an illustration, the schedule created by one participant is shown in Figure 1. (More examples of completed schedules can be found in Figure 4 or at https://osf.io/2smqj/.) Participants typically introduced a new topic every couple of days. Most participants arranged the four topics in the order of the arbitrary numerical labels (i.e., Topic 1, then Topic 2, and

Figure 1
Scheduling Task (Study 1)



Note. At the start of the task (top panel), participants were shown a blank 2-week schedule and colored boxes representing four lessons and 28 practice problems. Participants were instructed to move all lessons and problems into the schedule in any arrangement of their choosing. (Full instructions can be found at https://osf.io/2smqj/.) A completed schedule (example in bottom panel) included all lessons, problems, and 1 hr of unused class time. Unused time represented free time that the hypothetical students could spend on homework for other classes. (For other examples of completed schedules, see Figure 4.) See the online article for the color version of this figure.

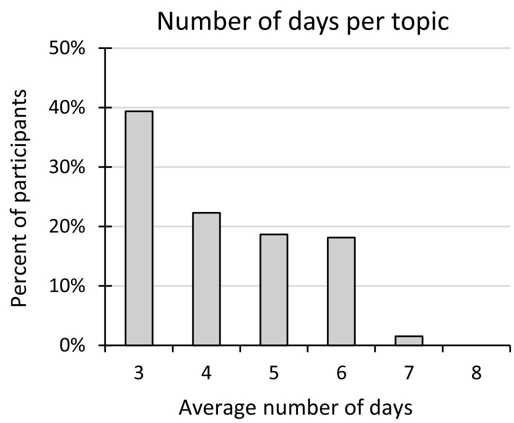
so forth), though this was not a requirement, and interchanging these labels would have no effect on the analyses reported below.

Degree of Spaced Practice

We measured the degree of spacing in each participant's schedule by counting the number of days (out of 10) that each topic appeared.[1] For instance, if exposures to Topic 3 (lesson or problem) occurred only during the Tuesday, Wednesday, and Friday of the second week, the number of days for that topic equaled three.

[1] While there are many possible ways to operationalize spacing and interleaving within a schedule (e.g., see Yan & Sana, 2021, for an alternative involving variance), we selected simple operationalizations that we believe are intuitive and practical for students and educators.

Figure 2
The Degree of Spacing in the Schedules Created by Participants (Study 1)



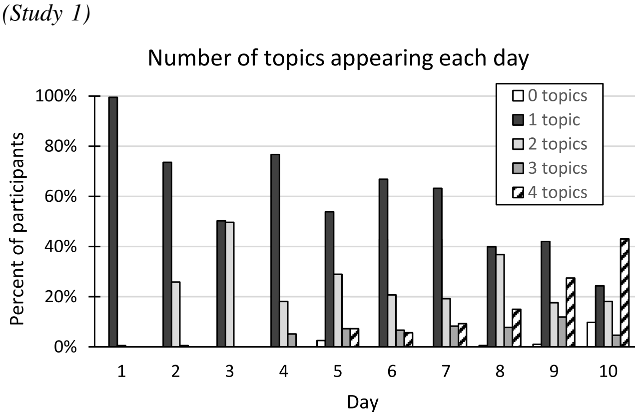
Note. The number of days that a topic appeared was averaged across topics and rounded to the nearest whole number in this figure. The task dictated that possible values ranged from 3 to 8 days. (N = 193)

For this measure, the minimum possible number of days was 3 days and the maximum possible was 8 days (see Figure 1). Averaged across topics, the mean number of days given to each topic was 4.1 (SE_M = 0.08, range = 3–7). Nearly 40% of participants scheduled topics to appear on the minimum possible number of days (Figure 2). Thus, participants showed a preference for the minimum amount of spacing permitted by the task.

Degree of Interleaved Practice

Most of the participants' schedules provided little interleaved practice. As an illustration, Figure 3 shows the number of topics appearing on each day of the 10-day schedules. Although the task allowed participants to schedule as many as four topics per day, the modal number of topics was one on every day except the last (Day 10). On Day 10, however, 43% of participants scheduled problems from all four topics, possibly because they believed the hypothetical students should review all topics immediately before the exam.

Figure 3
The Degree of Interleaving in the Schedules Created by Participants (Study 1)



Note. The figure shows the number of topics appearing during each day of the 10-day schedule. The mode was 1 for every day except Day 10.

We measured interleaving by determining the percentage of problems that were interleaved within class sessions. That is, a practice problem was counted as interleaved if its topic was different from the topic of the immediately preceding problem or lesson (e.g., a Topic 3 problem immediately following a Topic 1 problem). A problem was counted as blocked if its topic was the same as the immediately preceding problem or lesson (e.g., a Topic 3 problem immediately following another Topic 3 problem). A problem was excluded from the computation of this measure if it did not immediately follow another problem or lesson (i.e., those at the start of a class or following a blank space in the schedule). Averaged across participants, blocked problems were more than twice as common as interleaved problems (71% vs. 29%; $SE_M = 1.7\%$). In summary, when aiming to maximize learning for hypothetical students, most participants created schedules providing little spaced or interleaved practice.

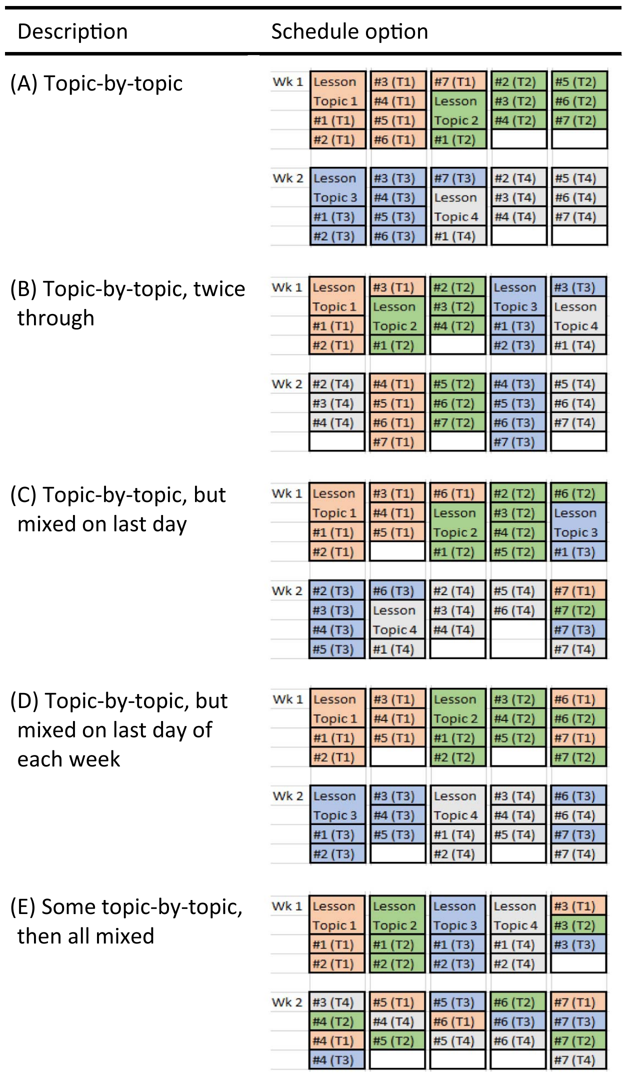
Interim Discussion

The results of the first study indicate that college students underappreciate the benefits of spaced and interleaved math practice. The results went beyond previous work by demonstrating this underappreciation in the context of math learning and by examining the quantity and timing of spaced and interleaved practice in the schedules participants created. When aiming to maximize learning for hypothetical students, most participants scheduled only a small dose of spacing and interleaving, including any mixed review during the last class before the exam. These results suggest that students' metacognitive knowledge might be improved by providing guidance about effective dosing (i.e., larger doses of spacing and interleaving) and timing (e.g., mixed practice should begin sooner, not just when an exam is imminent).

Still, some questions remain regarding students' underappreciation of these techniques. First, although we believe the scheduling task in Study 1 would allow participants to demonstrate their appreciation (if any) of these techniques, we wondered if modifications to the task might provide a more sensitive test of whether participants understand the utility of spaced or interleaved practice. For instance, perhaps participants could demonstrate better metacognitive knowledge if they were given ready-made schedule options rather than having to create their own schedules. This modification would allow for the possibility that participants might *recognize* good learning techniques even if they fail to generate the techniques on their own. Also, perhaps a longer test delay in the math scenario would lead participants to choose schedules with more spacing compared to a shorter test delay, since spacing is often less beneficial at short test delays (e.g., Bird, 2010; Rawson, 2012; Verkoeijen et al., 2008).

Second, what were the reasons behind students' scheduling decisions? Some theories of self-regulated learning suggest that students' understanding of when and why techniques are effective is an important part of metacognitive knowledge because it affects whether students can apply those techniques effectively and in appropriate circumstances (e.g., Butler & Winne, 1995; Pintrich et al., 2000; Pressley et al., 1989). To investigate these beliefs, we could ask participants to explain their rationale for their scheduling

Figure 4
The Five Schedule Options Provided to Participants (Study 2)



Note. Option descriptions (left column) were shown to participants alongside the schedule images. Option E (which has the most spacing and interleaving) would be expected to maximize test scores. See the online article for the color version of this figure.

decisions. We could also ask them to explicitly rate the utility of spacing and interleaving, which would reveal the relative strength of the beliefs. Finally, we could also investigate other beliefs (apart from beliefs about efficacy) that might help to explain why students neglect or resist schedules that are heavily spaced or interleaved. For example, do the schedules seem difficult, enjoyable, or representative of math courses they have experienced? Such questions could provide further insight into students' beliefs about spacing and interleaving and may suggest paths for improving their beliefs.

Study 2

In Study 2, we modified the scheduling task from Study 1 to give participants a better chance to demonstrate their appreciation of spacing and interleaving and to reveal the reasons underlying students' scheduling decisions. A new sample of participants received the same math scenario as in Study 1, but rather than create their own schedules, they answered each of several questions by selecting one schedule from among five options. The questions pertained to the perceived efficacy, difficulty, enjoyment, and representativeness of the schedules. Regarding efficacy, participants selected a schedule to maximize test scores in week 3 (a 3-day test delay, as in Study 1) and also selected a schedule to maximize test scores in week 7 (a 31-day test delay). We wondered whether participants' choices in Study 2 would be consistent with the schedules created in Study 1—or whether seeing schedule options might help participants recognize the utility of spacing and interleaving. We also wondered whether test delay would affect participants' scheduling choices. Following each scheduling choice, participants were asked to explain their rationale. Finally, participants also rated their agreement with several statements about the perceived value of spacing and interleaving.

Method

Participants

Participants were 175 undergraduates (135 women, 38 men, 2 unreported) recruited from the same participant pool as in Study 1. The sample size was large enough to give margins of error smaller than 7.5% for sample proportions (for schedule choices and reasons), smaller than 4% when measuring spacing or interleaving, and smaller than 5% for students' ratings of utility, at 95% confidence. Any student who had previously participated in Study 1 was ineligible for Study 2. Participants received credit in their psychology courses for research participation. The sample included 31% freshmen, 14% sophomores, 24% juniors, 30% seniors, and 1% other (e.g., nondegree seeking); the mean age was 21.1 years (SD = 5.2, range 18–51); and they self-identified as 73% White, 15% Black, 13% Asian, 1% other (2% unreported), and also 22% Hispanic (1% unreported).

Procedure

Participants completed the study online (approximately 12 min in total). The procedure consisted of two parts: A modification of the scheduling task from Study 1 and a survey of participants' beliefs about spacing and interleaving.

In the first part, participants were given the same scheduling scenario used in Study 1, but rather than creating their own unique schedules, participants selected from among five presented schedule options in response to several questions. The questions pertained to participants' perceptions of efficacy, difficulty, enjoyment, and representativeness, as well as open-ended questions about participants' rationale for their choices (Table 1). The five schedule options corresponded to five types of schedules commonly observed in Study 1 and had varying degrees of spaced and interleaved practice (Figure 4). The option with the most spacing and interleaving, option E, would be expected to maximize test scores. The schedule options were presented to participants in visual form, alongside short descriptions in layman's terms to help participants understand the features of each schedule.

Table 1
Questions About the Five Schedule Options (Study 2)

Questions
1. Which schedule do you think would be best for maximizing test performance on these topics on a *surprise test* that occurs *the following Monday* (week 3)?
2. Which schedule do you think would be best for maximizing test performance on these topics on a *surprise test* that occurs *a Month later* (week 7)?
3. Which schedule do you think students would find *most difficult* during the 2 weeks of learning and practice shown?
4. Which schedule would you *enjoy most* if you were a student in the class?
5. Which schedule is *most representative* of what you've typically experienced in math classes you've taken?

Note. Question 1 was most comparable to the task in Study 1. Questions 1 and 2 manipulated test delay, so we instructed participants to imagine a *surprise* test so they would not assume that hypothetical students would review immediately before a week 7 test. Immediately after responding to each of questions 1–4, participants were asked to explain why they selected the schedule they did.

In the second part of the study, participants rated their agreement (on a seven-point Likert scale from *Strongly Disagree* to *Strongly Agree*) for each of several statements (Table 2). Four items measured participants' opinions on the utility of spacing, massing, and interleaving practice problems.

Results and Discussion

Schedule Choices

Participants' schedule choices are shown in Figure 5. When aiming to maximize test scores in week 3, most participants (55%) favored options C and D, which provided only a small degree of spaced and interleaved practice. Fewer participants (18%) chose option E, which provided the most spacing and interleaving. When participants aimed to maximize week 7 test scores, the optimal schedule (option E) grew in popularity but did not surpass option D. Thus, test delay alone cannot fully explain why participants overlooked the benefits of spaced and interleaved practice. Furthermore, participants judged the optimal schedule (option E) to be most difficult, not especially enjoyable, and not representative of math classes they had taken. Participants indicated that option A (with the minimum possible spacing and no interleaving) was most representative of their math classes.

We also examined whether Study 2 choices were consistent with the results of Study 1. In the two studies, participants either designed a schedule to maximize test scores in week 3 (Study 1) or selected among existing schedules to maximize test scores in week 3 or week 7 (Study 2). These tasks pertained to the perceived efficacy of the schedules, and we compared the results across studies to examine the possibility that choices in Study 2 may have been affected by providing schedule options. Figure 6 shows, however, that the two studies were approximately consistent in terms of spacing (top half) and interleaving (bottom half). The two studies did not significantly differ in their average amounts of spacing, $t(366) = 1.8$, $p = .07$, $d = 0.19$, or interleaving, $t(366) = 0.9$, $p = .37$, $d = 0.09$, (80% power to detect effects as small as $d = 0.29$ with independent samples, two-tailed test, $\alpha = .05$). In sum, participants opted for little spacing and interleaving in both studies.

Following each schedule choice, participants were asked to explain the reasons for their choice. We identified a long list of reasons that appeared in the participants' responses, and since a single response could contain multiple reasons, we then coded each response for the presence or absence of each reason. In Table 3, we report reasons given by at least 10% of participants for the schedules they selected to maximize performance on week 3 or week 7 tests. For each test delay, around 45% of participants mentioned the importance of reviewing material, and 10%–15% emphasized that review can help to make information fresh in memory; these responses might suggest at least a rudimentary appreciation of spacing by some participants. Only 9% (not shown in table) touted the value of reexposure on separate occasions with time intervening—perhaps a more sophisticated understanding of spacing. Interestingly, 9%–13% of participants said they wanted their hypothetical students to see all topics or problem types close to the exam (i.e., on the last day of the schedule), which suggests these participants were seeking to minimize test delay rather than incorporate spacing or interleaving per se. Only about 5% of participants (not shown) mentioned benefits of interleaving such as helping to learn which strategy to apply; more commonly, participants endorsed arrangements that ran counter to interleaving (e.g., one topic at a time, 14%–20%; doing enough of a topic at once, 10%–15%). Finally, participants' explanations also revealed some of their concerns—about having sufficient clarity or enough time to process concepts—that could plausibly contribute to the underutilization of spacing and interleaving.

Ratings of the Utility of Spacing and Interleaving

Finally, participants were asked to consider the utility of spacing and interleaving by rating their agreement with the four statements shown in Table 2. Ratings indicated that opinions on spaced practice were not straightforward: Most participants agreed that spacing is beneficial (Statement 1), yet most participants also agreed that massing practice into a few focused assignments is better than more spaced assignments (Statement 2). One might expect that support for spacing (Statement 1) and support for massing (Statement 2) would be strongly negatively correlated, but they were not ($r = -.07$, $p = .36$). These seemingly contradictory beliefs highlight the value of investigating students' beliefs about competing techniques and dosage (discussed more below). Most notably, the value of interleaved practice was not consistently recognized (Statements 3 and 4). In fact, most participants did not see value in attempting problems of unknown type (Statement 4). Such attempts, however, allow students to practice identifying appropriate strategies and are a particularly beneficial feature of interleaved math practice.

General Discussion

In two studies presented here, we investigated college students' metacognitive knowledge of two highly effective learning techniques—spacing and interleaving—in the context of math learning. We used a scheduling task that enabled participants to communicate concrete, visual depictions of learning schedules they believed to be most effective. As important, the task enabled us to examine the timing and quantity of spacing and interleaving in those schedules. We also investigated participants' reasons for their scheduling choices. In short, we investigated a combination of

Table 2

Participants' Ratings of the Utility of Spaced Practice, Massed Practice, and Interleaving (Study 2)

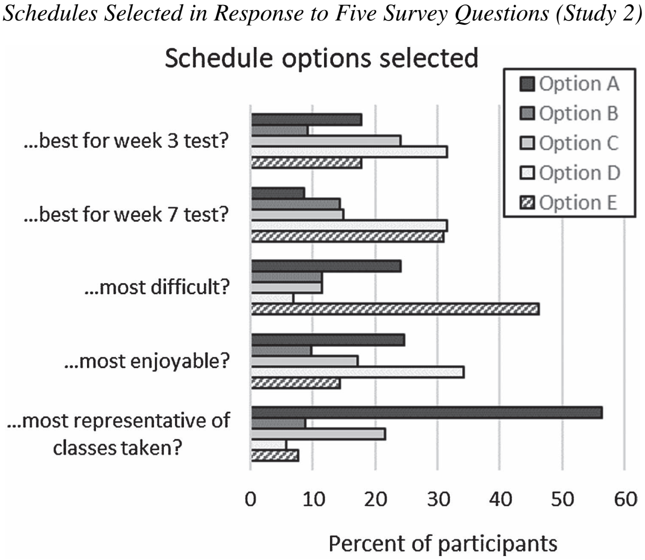
Statements	About the utility of ...	Mean rating	Percent of participants who selected ratings from 1 (*strongly disagree*) to 7 (*strongly agree*)						
			1	2	3	4	5	6	7
When students learn a particular math concept (or topic), it's important for practice problems on that concept to be spread across many days.	Spaced practice	6.0*	0	1.7	2.9	4.0	19.4	30.3	41.7
If a student will practice a total of seven problems on Concept A, it's better for their learning if they practice those problems (on Concept A) in one or two focused assignments rather than spread them across many assignments.	Massed practice	4.6*	1.1	12.6	20.6	10.9	18.3	18.9	17.7
When doing math practice problems, it's best for students' learning if they practice a variety of topics within a single practice session.	Interleaving (mix)	3.9	10.9	16.6	20.0	6.3	25.7	13.1	7.4
When doing math practice problems, it's good for students' learning if they don't know what topic will be the focus of the next problem.	Interleaving (unknown focus)	3.4*	17.1	25.1	14.9	10.3	14.3	10.3	8.0

Note. A rating of 4 (midpoint) represented *neither agree nor disagree*. Mean ratings are marked with asterisks if they differed from the midpoint ($p < .001$; 80% power to detect effects as small as $d = 0.21$ with two-tailed, one-sample *t*-tests, $\alpha = .05$, $N = 175$). Ratings for statements 3 and 4 (about interleaving) were positively correlated ($r = .31$, $p < .01$); all other pairs were unrelated.

the declarative, procedural, and conditional knowledge (or beliefs) that constitute metacognitive knowledge (Pintrich, 2000) to give a more detailed description than previously reported of college students' beliefs about spacing and interleaving. We found that college students underappreciated the utility of these techniques for math learning, and we also identified specific flaws in their beliefs.

Figure 5

Schedules Selected in Response to Five Survey Questions (Study 2)



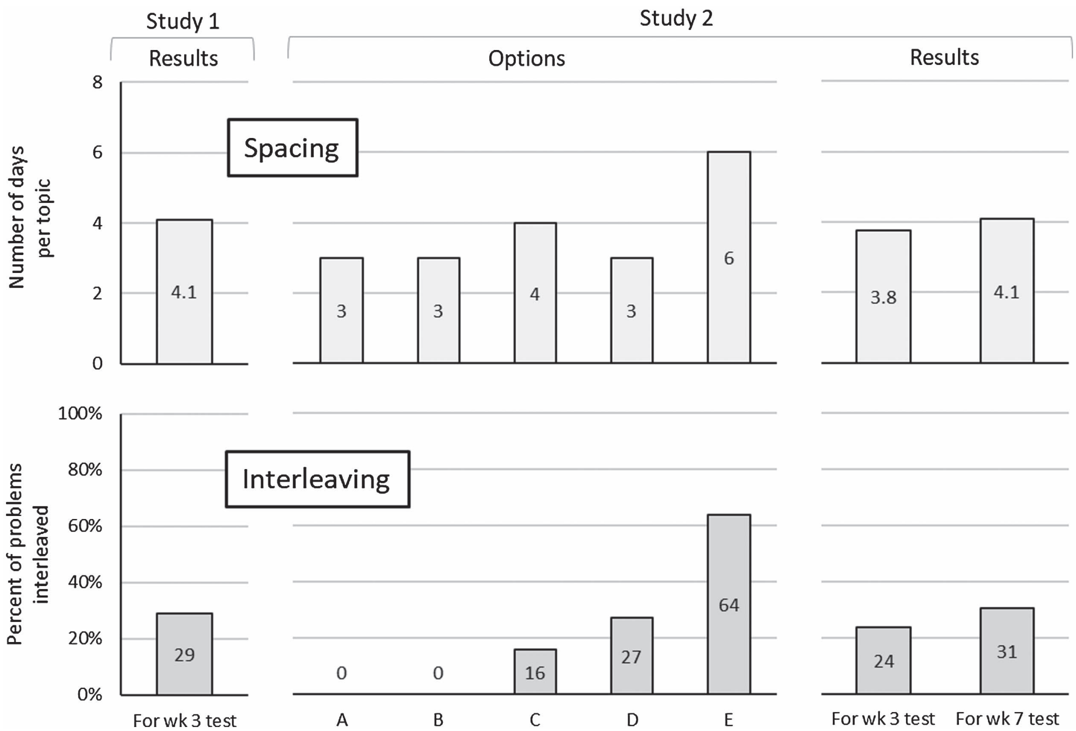
Note. The survey questions are abbreviated here; full questions are provided in Table 1. Each option is shown in Figure 4. For the two questions relating to test score, a large majority of participants did not choose the empirically supported techniques (option E).

Prior research has shown an underappreciation of spacing and interleaving in other contexts (e.g., Cohen et al., 2013; Yan et al., 2016), but it is particularly noteworthy for mathematics. College students have considerable familiarity with math learning and have undoubtedly experienced spacing of math concepts and interleaving of math problems, at least occasionally, during their years of schooling. This experience might be expected to promote metacognitive knowledge of these techniques (Butler & Winne, 1995; Pintrich et al., 2000; Pressley et al., 1989; Schneider, 2008). Furthermore, math courses commonly feature lots of practice problems that are suitable for spaced and interleaved practice. Thus, mathematics provided a favorable context for evaluating students' awareness of these techniques, compared to less familiar domains or less suitable materials. Nonetheless, the familiarity of math learning and the suitability of math problems did not ensure that students possessed knowledge of spacing or interleaving or that their beliefs about optimal scheduling were accurate. Rather, many students neglected these techniques and displayed errors in their beliefs—which we suspect may extend to other domains as well, including other quantitative or problem-focused domains that would also benefit from spaced and interleaved practice.

Why did prior experience with math learning not produce good metacognitive knowledge of spacing and interleaving? Though many reasons are plausible, one reason might be that most students' prior math experience involved low levels of these techniques. Indeed, our sample of college students reported that their previous math classes did not typically utilize much spacing and interleaving (Study 2). Thus, students might erroneously infer that low levels of these techniques are the ideal amount. More generally, if prior experience predominantly involves inferior learning techniques,

Figure 6

Spacing and Interleaving in the Schedules Created by Participants (Study 1) or Chosen by Participants (Study 2)



Note. The upper panel shows the amount of spacing, which was defined as the number of days per topic (minimum possible = 3, maximum possible = 8). The bottom panel shows the percent of practice problems that were interleaved. Study 1 results are shown on the left, and Study 2 results are shown on the right. In the middle, for comparison, we display the amount of spacing and interleaving in each schedule option of Study 2.

then the utility of superior techniques may go unrecognized. Further, even if a technique is amply experienced, recognition of its utility may depend on whether students monitor or receive feedback about the technique or its alternative (see Butler & Winne, 1995). Yet another plausible reason is that any prior experience with spacing or interleaving in their classes may have limited impact on metacognitive knowledge if students do not understand the decision-making involved—that is, why the techniques were used or when they are most appropriate. As mentioned previously, this kind of conditional knowledge helps learners to apply learning techniques effectively and in appropriate circumstances.

The present studies allowed us to identify specific gaps or errors in students' metacognitive knowledge about spacing and interleaving. For instance, in both studies, most participants mistakenly

Table 3

Most Common Reasons Given by Participants for the Schedules They Selected to Maximize Test Scores (Study 2)

Reasons for schedule choices	Percent of participants who listed the reason	
	For week 3 test (%)	For week 7 test (%)
Good to review accumulated knowledge	46	45
Helps memory/reduces forgetting	21	29
Should grasp topic before moving to next topic/one-by-one	20	14
Should learn a topic and focus on it/practice enough at once	15	10
Helpful for prepping for test	14	17
Reduces confusion/increases clarity or understanding	14	15
Gives time to learn/process/comprehend topics or problems	13	10
Should be exposed to all topics or problem types close to exam	13	9
Review makes information refreshed in memory	10	15

Note. Percentages do not sum to 100% because participants could cite multiple reasons. Table shows only those reasons cited by at least 10% of participants (for either week 3 or 7).

believed that performance would be maximized with only a small dose (if any) of spacing and interleaving. That is, most students created (Study 1) or chose (Study 2) schedules with minimal spacing and interleaving—contrary to more heavily spaced and interleaved schedules that would maximize performance. These small doses of spacing and interleaving were often driven by the activities scheduled for the last class meeting before the exam—when many participants scheduled a mix of problems to serve as a review. Review that occurs shortly before exams is common in classrooms, and such review does increase the amount of spacing and interleaving. However, in many classes, exams occur infrequently, so if students rely on exam review to be their main source of spaced and interleaved practice, they will receive much less spacing and interleaving than would be optimal for learning. Further, based on the rationale participants provided, our results suggest that choosing to review did not usually signify a sophisticated understanding of the benefits of spacing across days or interleaving within sessions but instead may indicate their wish to refresh memory on each topic close to the test. In other words, they are shortening the retention interval before the test, which is independent of spacing and interleaving.

Importantly, even though Study 2 provided ready-made schedule options, participants still favored smaller, rather than larger, amounts of spacing and interleaving when trying to optimize test scores. Thus, seeing a schedule option with heavy spacing and interleaving did not help participants to recognize its utility. Study 2 also allowed for the possibility that participants in Study 1 undervalued spacing and interleaving because these techniques are less advantageous at shorter test delays (like that in Study 1). However, while participants slightly altered their choices based on test delay (perhaps because the presence of both questions implied they should), they did not favor heavy spacing and interleaving at either delay. Thus, most participants showed little awareness that spacing is especially advantageous when aiming to retain knowledge across long test delays.

We also identified a variety of possible reasons for college students' resistance to spacing and interleaving (Study 2). For instance, in participants' rationales for their scheduling choices, they expressed having concerns about clarity, wanting sufficient processing time, and thinking linearly about topics—that is, perspectives that may contribute to minimal use of spacing and interleaving. Also, by directly querying participants' perceptions of difficulty and enjoyment of the schedule options, we learned that the schedule option with the most spacing and interleaving was perceived to have high difficulty and low enjoyment, which may discourage its use. Indeed, students may not recognize that the difficulty of spaced and interleaved practice is a desirable difficulty, yielding benefits to future performance (Bjork & Bjork, 2011). Instead, they might conclude that the difficulty of a technique signals low efficacy (Kirk-Johnson et al., 2019).

Unfortunately, failing to recognize effective learning techniques, including spacing and interleaving, is detrimental for students—particularly in the context of self-regulated learning (when students have much control over their practice), which is common in college-level or adult education (for further discussion, see Kornell & Finn, 2016). Underappreciation of effective techniques is not limited to students, however. Many teachers (Halamish, 2018; Morehead et al., 2016) and textbooks (Rohrer, Dedrick, & Hartwig, 2020) also provide little spaced and interleaved practice, which is consistent with participants' claims (in Study 2) that their previous math classes have not typically utilized much spacing and interleaving. Interestingly, when selecting schedules for their hypothetical students, participants did not simply select the blocked schedule most representative of their previous math classes, but they nevertheless chose low levels of spacing and interleaving. Future research might examine the extent to which improving teachers' knowledge of these techniques yields improvements in students' metacognitive knowledge.

Finally, we also asked participants to explicitly rate the utility of spacing and interleaving, which further clarified the shortcomings of their beliefs. For instance, their ratings of interleaving indicated that participants were generally unaware of the advantages of mixed practice or practicing problems of unknown type. Also, their ratings of spacing revealed seemingly contradictory beliefs about the advantages of spacing versus massing. That is, when participants rated the utility of spacing, most agreed that spacing is valuable; yet they also agreed that focused assignments are better than spacing and incorporated minimal spacing into their schedules. Previous research similarly suggests that the utility of spacing is sometimes recognized (e.g., Susser & McCabe, 2013) and sometimes not (e.g., Taraban et al., 1999). These results may seem contradictory, but we believe they highlight the importance of understanding the nuances of students' beliefs about learning techniques—including their beliefs about competing alternatives. In other words, participants assigned some value to spacing, but they also assigned value to the alternative (massing), which likely contributes to underutilization of spacing. In sum, to understand students' metacognitive knowledge about learning techniques, researchers must go beyond simple awareness of utility and, as we do here, also consider students' beliefs about how, when, and why the techniques (or their alternatives) should be used.

Limitations

A possible limitation of these studies pertains to external validity. Whether our results can be generalized beyond our college student sample is unknown, and future research could use the present methods with other groups of interest, such as undergraduate preservice teachers or practicing K-12 teachers. Also, our task instructions did not specify characteristics of the hypothetical students nor which math course participants should imagine (e.g., high school algebra), but future research may find such specification informative. Further, our scheduling task, like any scheduling task or scenario, had somewhat arbitrary features. For example, the schedules were limited to a 2-week period, four topics, and hour-long class periods. Longer scheduling periods, as well as other test delays, different numbers of topics or problems, and different lengths of lessons, activities, or class periods, might produce different results. A related caveat is that our scheduling task required each topic to be seen on at least 3 days, at minimum. Many participants opted for the minimum amount of spacing, but we do not know whether they would have spaced less if permitted. Previous (non-math) scheduling studies, which allowed participants to choose the days on which studying occurred (rather than daily class time), indicate that at least some participants might opt for fewer than 3 days (e.g., Wissman et al., 2012). Still, we believe the classroom scenario used here is ecologically valid because, in many classes, fully massing a topic within a single class period is not feasible.

A second limitation, relevant to Study 1, is that participants may have designed schedules with mostly massed and blocked practice because those schedules are the easiest to design or are considered to be the default arrangement. Perhaps many teachers and textbooks favor massed and blocked practice for these same reasons. However, the results of Study 2 help to alleviate this concern, because the ready-made options enabled participants to select more complex schedules with no extra effort, yet schedules with little spacing and interleaving were still preferred.

A third limitation, relevant to Study 2, is that the set of options a researcher provides can influence which options participants find most appealing. Perhaps other variations of schedules with substantial spacing and interleaving could be more attractive to participants. Also, schedule features unrelated to spacing and interleaving, such as the placement of unscheduled free time, might increase or decrease the appeal of an option. Even so, the consistency of results across these two studies shows that the set of options provided in Study 2 did not radically alter participants' responses.

Implications

Spacing and interleaving are learning techniques that have broad applicability, yet students' beliefs about these techniques are faulty. In the present study, the utility of spaced and interleaved math practice was not intuitive to college students, despite their experience with math learning and the suitability of the materials. The accuracy of students' beliefs about effective learning is important because beliefs can shape behavior. That said, beliefs about optimal study do not solely determine behavior; many other factors (e.g., enjoyment of material, juggling of many academic demands) also affect behavior and may even produce behavior that the learner does not believe to be optimal. Nonetheless, to the extent that students' beliefs about learning are inaccurate, one possible lever to help improve student study behavior is education that increases students' knowledge of effective learning techniques and dispels misconceptions about learning.

The present study identified specific shortcomings in students' metacognitive knowledge of spacing and interleaving with respect to perceived utility, implementation (e.g., dose and timing), and students' rationales. The shortcomings have implications for how students' beliefs might be improved. For instance, the present sample of students would probably benefit relatively little from simply being taught that "spacing is an effective technique"—because most already agree and might already choose a small amount of spacing and interleaving by reviewing before an exam. Indeed, many participants did embrace the idea of reviewing before a test, yet they did not recognize the greater utility of spacing practice across a larger number of days or starting to interleave problem types sooner. The participants might benefit, instead, from being taught to use a higher dose of these techniques spread across time (not only before exams). They should also be aware that the difficulty associated with using these techniques is beneficial to their learning—not a sign of low efficacy. Further, the participants would likely benefit from understanding when and why these techniques are more effective than alternatives. By identifying faulty beliefs or gaps in students' metacognitive knowledge, as done here, research can point to features of study skills training that could serve students best.

References

Barzagar Nazari, K., & Ebersbach, M. (2019). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Applied Cognitive Psychology*, *33*(2), 288–298. https://doi.org/10.1002/acp.3485

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*(4), 635–650. https://doi.org/10.1017/S0142716410000172

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory*, *25*(6), 784–792. https://doi.org/10.1080/09658211.2016.1221974

Borkowski, J. G., Chan, L. K. S., & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. In G. Schraw & J. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 1–41). Buros Institute of Mental Measurements.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Carpenter, S. K. (2014). Spacing and interleaving of study and practice. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying the science of learning in education: Infusing psychological science into the curriculum* (pp. 131–141). American Psychological Association.

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*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cohen, M. S., Yan, V. X., Halamish, V., & Bjork, R. A. (2013). Do students think that difficult or valuable materials should be restudied sooner rather than later? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1682–1696. https://doi.org/10.1037/a0032425

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, *53*, 63–147. https://doi.org/10.1016/S0079-7421(10)53003-2

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*(4), 309–330. https://doi.org/10.1007/BF01320097

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Eglington, L. G., & Kang, S. H. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 475–485. https://doi.org/10.1016/j.jarmac.2017.07.005

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition*, *47*(6), 1088–1101. https://doi.org/10.3758/s13421-019-00918-4

Halamish, V. (2018). Pre-service and in-service teachers' metacognitive knowledge of learning strategies. *Frontiers in Psychology*, *9*, Article 2152. https://doi.org/10.3389/fpsyg.2018.02152

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*(1), 126–134. https://doi.org/10.3758/s13423-011-0181-y

Helsdingen, A., van Gog, T., & van Merriënboer, J. (2011a). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology*, *103*(2), 383–398. https://doi.org/10.1037/a0022370

Helsdingen, A. S., van Gog, T., & van Merriënboer, J. J. (2011b). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*, *21*(1), 126–136. https://doi.org/10.1016/j.learninstruc.2009.12.001

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *28*(4), 853–873. https://doi.org/10.1007/s10648-015-9349-8

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 12–19. https://doi.org/10.1177/2372732215624708

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, Article 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219–224. https://doi.org/10.3758/BF03194055

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*(2), 498–503. https://doi.org/10.1037/a0017807

Kornell, N., & Finn, B. (2016). Self-regulated learning: An overview of theory and data. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 325–340). Oxford University Press.

Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. S. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *32*(1), 277–295. https://doi.org/10.1007/s10648-019-09489-x

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, *35*(2), 105–123. https://doi.org/10.1901/jaba.2002.35-105

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*(3), 462–476. https://doi.org/10.3758/s13421-010-0035-2

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, *24*(2), 257–271. https://doi.org/10.1080/09658211.2014.1001992

Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (2019). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology*, *111*(7), 1172–1188. https://doi.org/10.1037/edu0000336

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Buros Institute of Mental Measurements.

Pressley, M., Borkwski, J. G., & Schneider, W. (1989). Good information processing: What it is and how education can promote it. *International Journal of Educational Research*, *13*(8), 857–867. https://doi.org/10.1016/0883-0355(89)90069-4

Pyc, M. A., & Dunlosky, J. (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory & Cognition*, *38*(4), 431–440. https://doi.org/10.3758/MC.38.4.431

Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction*, *23*, 98–114. https://doi.org/10.1016/j.learninstruc.2012.07.003

Rawson, K. A. (2012). Why do rereading lag effects depend on test delay? *Journal of Memory and Language*, *66*(4), 870–884. https://doi.org/10.1016/j.jml.2012.03.004

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, *25*(4), 523–548. https://doi.org/10.1007/s10648-013-9240-4

Rawson, K. A., Vaughn, K. E., Walsh, M., & Dunlosky, J. (2018). Investigating and explaining the effects of successive relearning on long-term retention. *Journal of Experimental Psychology: Applied*, *24*(1), 57–71. https://doi.org/10.1037/xap0000146

Roediger, H. L., III, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

Rohrer, D., Dedrick, R. F., & Hartwig, M. K. (2020). The scarcity of interleaved practice in mathematics textbooks. *Educational Psychology Review*, *32*(3), 873–883. https://doi.org/10.1007/s10648-020-09516-2

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C.-N. (2020). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology*, *112*(1), 40–52. https://doi.org/10.1037/edu0000367

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*(1), 84–98. https://doi.org/10.1037/edu0000119

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain and Education : the Official Journal of the International Mind, Brain, and Education Society*, *2*(3), 114–121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schutte, G. M., Duhon, G. J., Solomon, B. G., Poncy, B. C., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *Journal of School Psychology*, *53*(2), 149–159. https://doi.org/10.1016/j.jsp.2014.12.003

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, *19*(1), 107–122. https://doi.org/10.1002/acp.1066

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*(5), 763–767. https://doi.org/10.1002/acp.1747

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244–253. https://doi.org/10.3758/MC.38.2.244

Susser, J. A., & McCabe, J. (2013). From the lab to the dorm room: Metacognitive awareness and use of spaced study. *Instructional Science*, *41*(2), 345–363. https://doi.org/10.1007/s11251-012-9231-8

Taraban, R., Maki, W. S., & Rynearson, K. (1999). Measuring study time distributions: Implications for designing computer-based courses. *Behavior Research Methods, Instruments, & Computers*, *31*(2), 263–269. https://doi.org/10.3758/BF03207718

Taylor, K., & Rohrer, D. (2010). The effect of interleaving practice. *Applied Cognitive Psychology*, *24*(6), 837–848. https://doi.org/10.1002/acp.1598

Toppino, T. C., & Cohen, M. S. (2010). Metacognitive control and spaced practice: Clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1480–1491. https://doi.org/10.1037/a0020949

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Özsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology*, *22*(5), 685–695. https://doi.org/10.1002/acp.1388

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Erlbaum.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory*, *20*(6), 568–579. https://doi.org/10.1080/09658211.2012.687052

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177

Yan, V. X., & Sana, F. (2021). Does the interleaving effect extend to unrelated concepts? Learners' beliefs versus empirical evidence. *Journal of Educational Psychology*, *113*(1), 125–137. https://doi.org/10.1037/edu0000470

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied*, *23*(4), 403–416. https://doi.org/10.1037/xap0000139

Received September 16, 2020
Revision received June 26, 2021
Accepted July 12, 2021 ■