Running head: STRONG EFFECTS ON WEAK THEORETICAL GROUNDS

Forthcoming in *Zeitschrift für Psychologie*, 2014, Vol 222 (2)
http://www.hogrefe.com/periodicals/zeitschrift-fuer-psychologie/

Strong Effects on Weak Theoretical Grounds: Understanding the Distributed Practice Effect

Carolina E. Küpper-Tetzel

Center for Integrative Research on Cognition, Learning, and Education (CIRCLE)
Washington University in St. Louis, USA

Corresponding author's address:
Carolina E. Küpper-Tetzel
Center for Integrative Research on Cognition, Learning, and Education (CIRCLE)
Washington University in St. Louis
One Brookings Drive
St. Louis, MO 63130
United States
Phone: 314 935-6142
Fax: 314 935-9455
Email: carolina.kuepper-tetzel@wustl.edu

Abstract

The distributed practice effect is one of the most researched memory effects in cognitive psychology. Beneficial distributed practice effects for long-term retention have been demonstrated in different domains and they are remarkably large in size, too. However, despite strong effects, this research field still lacks a unified theory offering explanations for a wide range of findings. This article reviews empirical studies on the distributed practice effect that have immediate relevance for educational settings. Against the backdrop of this review, the article discusses theory candidates and ways of specifying them for empirical tests using nonstandard statistical methods. I conclude that future studies will have to fine-tune theories to strengthen the significance of empirical results and to allow for better recommendations to educators. This promises to increase the enthusiasm to systematically implement distributed practice in instruction routines and bridge psychological research and educational practice.

Keywords: Distributed practice; Theories; Educational Practice; Applied Memory Research; Review

Strong Effects on Weak Theoretical Grounds: Understanding the Distributed Practice Effect

Back in 1885 Hermann Ebbinghaus conducted the first studies on the advantage of distributed practice over massed practice with himself as the only participant. His experimental method and findings were groundbreaking and had a great influence on successive research on memory and learning. Ebbinghaus (1885/1964) observed that to achieve the same learning outcome, fewer repetition trials were needed when learning sessions were distributed across time compared to when all learning occurred crammed in a single day. To put it in his words: "It makes the assumption probable that with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time" (p. 89). As a consequence, hundreds of studies have been conducted to examine this learning effect (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

The simplest research design to investigate the distributed practice effect is displayed in Figure 1 and consists of two learning sessions separated by an interstudy interval (ISI). Participants are introduced to the to-be-learned material during the initial learning session and review the same material during a relearning session after a specific ISI, which is usually manipulated. Memory is tested in the final test session after the retention interval (RI). The RI is measured from the last learning session to the final session. The RI can be fixed or varied, too. Importantly, no additional learning takes place during the RI. Adding relearning sessions is a straightforward way to bring the research design closer to more authentic settings in which the same material is usually revisited on more than one occasion. Generally, the literature distinguishes the *spacing effect*, which describes the comparison between spaced practice (i.e., ISI > 0) and massed practice (i.e., ISI = 0), and the *lag effect*, which is the comparison of different non-zero ISIs between learning sessions. In accordance with Cepeda et al. (2006), I use the term *distributed practice* to embrace both effects.

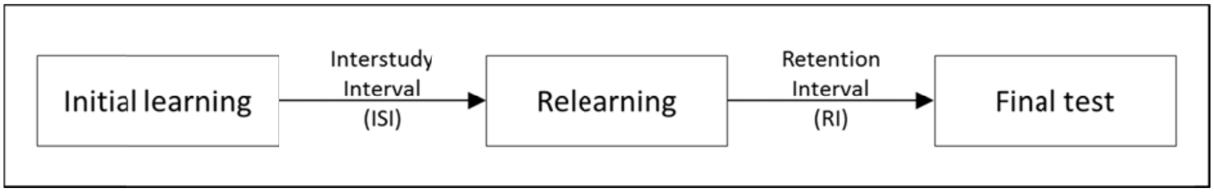

Figure 1. Visualization of the simplest research design of the distributed practice effect. It comprises one initial learning session and a relearning session separated by a specific ISI. Final memory performance is measured after the RI.

Furthermore, two research paradigms have to be discriminated: the *within-session paradigm* and the *between-sessions paradigm*. In the past, those paradigms have often been treated as interchangeable, which is problematic for theory development. For the *within-session paradigm,* an item is part of a list of items and is relearned within this list. Distributed practice is realized by learning of other (intervening) items that are interspersed between initial presentation and repeated presentation of an item. Thus, the ISI is usually quite short in the range of a couple of seconds or minutes at most. For the *between-sessions paradigm*, the to-be-learned material is initially studied once and then restudied after a specific ISI. This paradigm allows investigating varying ISIs between learning sessions in the range of days, months, or even longer periods, resulting in greater proximity to and significance for real-world learning than within-session studies. Notwithstanding, the latter paradigm may inform practice, too, as it has implications for whether students should practice the same problem or item in a blocked or interleaved fashion, where different problems or items alternate (e.g., Helsdingen, van Gog, & van Merriënboer, 2011; Mayfield & Chase, 2002; Taylor & Rohrer, 2010). Delaney, Verkoeijen, and Spirgel (2010) provided an excellent review on studies that used the within-session paradigm. Apart from presenting important findings, they offer an in-depth analysis of theoretical accounts for empirical findings. In contrast, this work focuses on between-session designs when reviewing the literature and evaluating theories.

Overview

The distributed practice effect has mainly been shown for verbal learning tasks, such as paired associates learning (e.g., Cepeda et al., 2009; Glenberg & Lehmann, 1980) or learning of lengthy text passages (e.g., Ausubel, 1966; Rawson & Kintsch, 2005), but also – albeit to a lesser extent – for learning in mathematics (e.g., Gay, 1973; Rohrer & Taylor, 2007) and natural sciences (e.g., Reynolds & Glaser, 1964; Vlach & Sandhofer, 2012). Moreover, beneficial effects of distributed practice have been revealed for the learning of simple (e.g., Shea, Lai, Black, & Park, 2000) as well as educationally relevant complex motor skills (e.g., music: Simmons, 2012, sports: Lawton, Cronin, & Lindsell, 2006, medicine: Moulton et al., 2006).

Among researchers there is no doubt that the distributed practice effect can reduce forgetting and promote long-term retention in very different educationally relevant domains. Besides a few exceptions (e.g., Carpenter, Pashler, & Cepeda, 2009; Küpper-Tetzel, Erdfelder, & Dickhäuser, 2013; Sobel, Cepeda, & Kapler, 2011), the distributed practice effect has been examined mainly in the laboratory. Generally, effects are remarkably large in size, ranging from Cohen's $d = 0.71$ (Hattie, 2009) to Cohen's $d > 1$ (Cepeda et al., 2006). This is good news for the application to real-world settings because it means that benefits of distributed practice will probably hold against added variance introduced by naturalistic environments. In addition, the distributed practice effect is an easy-to-implement (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012) and cost-effective (Roediger & Pyc, 2012) instruction and learning strategy. It can be introduced by the instructor in class, but also adopted by the learner during self-regulated learning.

Despite these strong findings, researchers report that a transition of promising cognitive principles such as the distributed practice effect to real-world settings has not yet happened. The large number of articles in scientific journals (e.g., Carpenter et al., 2012; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Pyc, 2012) as well as articles in newspapers or online blogs (e.g., Castillo, 2013, January 10; Paul, 2011, September 2010) testify to the missing bridge between research and practice.

Dempster (1988) pointed out nine reasons for why implementing the distributed practice effect in education has failed. Most of his concerns have been examined in different studies, e.g. demonstration of distributed practice in educational settings, use of educationally relevant material. However, one crucial point has, to date, not been tackled satisfactorily: the *lack of a valid theory promoting understanding of the distributed practice effect*. Although we are dealing with a very old learning phenomenon, we have not gotten to the bottom of its mechanisms. I agree with Dempster when he argues that “our theoretical ignorance may have been and may continue to be an impediment to application or might contribute to inappropriate applications” (p. 632). I am convinced that putting the distributed practice effect on strong theoretical grounds is a vital component for a better communication between research and practice. In contrast to previous reviews (e.g., Carpenter et al., 2012; Cepeda et al., 2006), this review draws attention to this often trivialized aspect, emphasizes the importance of theories, and connects empirical findings to theory.

In the following, I first review empirical findings on beneficial distributed practice effects in the domains of rote learning of verbal material and conceptual learning in mathematics and science as they have direct implications for real-world practice. Then, I elaborate on theoretical explanations of the distributed practice effect, highlight their validity for the empirical findings, point to their weaknesses, and propose an alternative way of testing them. Finally, I conclude by calling attention to the importance of better understanding the distributed practice effect for facilitating the implementation in real-world practice.

Distributed Practice Effect in Verbal Learning

The majority of empirical findings suggests that benefits of spaced practice are particularly pronounced on delayed memory tests, but often absent on tests immediately following practice (for exceptions see Glover & Corkill, 1987; Krug, Davis, & Glover, 1990). This insight was already noticed by Austin (1921), who concluded that “it is for retention after longer intervals that the value of divided repetitions is particularly noticeable” (p. 390). The interaction between practice and RI explains why students often stick to a *cramming* strategy as it has resulted in good performances on immediate exams. An early study by Gordon (1925), for example, revealed that participants performed better on an immediate free recall test when a text passage was read out repeatedly in immediate succession compared to when practice was divided into two sessions, separated by a 3-day ISI. However, after an RI of 4 weeks, text memory was clearly better for spaced than for massed practice. Echoing this finding, a more recent study by Rawson and Kintsch (2005) had participants reread text passages either in a massed fashion or after a 1-week ISI. Memory performance on an immediate free recall test was better when the text passage was studied in a massed fashion, while the spaced rereading condition was not different from a single exposure condition. On a delayed test 2 days later, however, free recall performance clearly benefited from spaced rereading, and the massed rereading condition dropped to the level of the single exposure condition. Beneficial spacing effects on delayed tests have been confirmed in a classroom study as well (Sobel, Cepeda, & Kapler, 2011). They taught fifth graders Graduate Record Examinations (GRE) vocabulary words and their meaning in a massed or spaced fashion (7-day ISI) using a test-with-feedback procedure. On a test 5 weeks later, they found superior memory performance for words that were practiced in a spaced manner, translating in a memory improvement of 177% compared to massed practice.

An important issue for real-world learning revolves around the question how ISIs of different lengths affect retention of verbal material (*lag effect*). Empirical findings suggest that there is a systematic relationship between ISI and RI, which is far more complex than “long ISIs are better than short ISIs.” An early study by Ausubel (1966), for example, found that participants who reread a text passage with a 1-day ISI performed descriptively better on a final test 6 days later than participants who reread it with a 7-day ISI. Verkoeijen, Rikers, and Özsoy (2008) had participants read and reread a text passage either in immediate succession, separated by 4 days, or separated by 3.5 weeks. Performance on a free recall and a short-answer test after a fixed RI of 2 days for all ISI conditions increased between the massed and the 4-day ISI condition, but decreased for the 3.5-week ISI condition. Interestingly, memory performance did not differ between the massed condition and the 3.5-week ISI condition. Thus, longer ISIs may not by all means have a positive effect on performance measured on a later memory test. Glenberg and Lehmann (1980) found similar intriguing results using an incidental learning task in which participants studied word pairs. Again, when retention was assessed 7 days after the last learning episode, memory was better when learning sessions had been separated by a 1-day ISI instead of a 7-day ISI. To investigate the relationship between ISI and RI in more detail, Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) combined RIs ranging from 7 days to 350 days and ISIs ranging from 0 to 105 days in a web-based experiment. Participants were asked to learn and relearn unknown trivia facts through test-with-feedback trials (e.g., *What European nation consumes the spiciest Mexican food? Norway.*). Their results showed that the optimal time for relearning depends on the length of the RI. For each RI, memory performance followed a nonmonotonic trend by first increasing with ISI until reaching an optimal ISI and then decreasing again. The ratio of optimal ISI to RI decreased with increasing RIs. For instance, recall performance for a 1 week RI was best when the to-be-learned material was practiced with an ISI of 1 day. ISIs shorter or longer than 1 day led to detrimental memory performance. However, for an RI of 35 days, recall performance increased up to an ISI of 11 days and decreased for longer ISIs.

The inverted-U-shaped trend seems to be a reliable finding which is robust against changes in material (e.g., expository texts: Verkoeijen et al., 2008, question-answer pairs: Cepeda et al., 2009, or word pairs: Küpper-Tetzel & Erdfelder, 2012), but also against changes in setting and population as Küpper-Tetzel, Erdfelder, and Dickhäuser (2013) demonstrated. They conducted an applied study in a German middle school and had sixth graders study and restudy foreign vocabulary pairs using a test-with-feedback procedure during their regular English class, with ISIs varying from 0, 1, to 10 days. Students were tested with a free and cued recall test 7 or 35 days later. Consistent with earlier research, Küpper-Tetzel et al.'s results show that students' memory for vocabulary assessed after 7 days benefited the most from a 1-day ISI, producing a 35% and 34% increase in performance compared to 0-day and 10-day ISI, respectively, with Cohen's *d*s ≥ 1.69. With an RI of 35 days, however, memory performance improved for an ISI of up to 10 days, resulting in a 28% improvement compared to 0-day ISI for Cohen's $d = .87$.

In summary, studies on verbal learning using two learning sessions demonstrate that memory performance on a final test after educationally relevant RIs benefits from distributed practice. Even though most of these studies have been conducted in the laboratory, few applied studies complement and corroborate the main findings on the beneficial effects of distributed practice (Bird, 2010; Carpenter et al., 2009; Küpper-Tetzel et al., 2013; Sobel et al., 2011). It bears mentioning though that the relationship between ISI and RI is more complex, so that a rule like "the longer the ISI, the better memory performance" is too simplistic. The systematic interaction cautions that choosing a too short or too long ISI for given RIs may harm performance on a future test.

Although investigating the optimal distribution of two learning sessions has some merit – especially in view of severe time constraints that instructors face –, students and teachers will most probably engage in more than one relearning session for reviewing past material. For them, it would be important to know how multiple relearning sessions should be optimally scheduled across time to improve maintenance of knowledge.

For research designs with more than two relearning sessions, three schedule formats can be distinguished: ISIs between learning sessions are either constant over time (equal schedule; X--X--X), increase over time (expanding schedule; X-X---X), or decrease over time (contracting schedule; X---X-X). For an expanding schedule the ISI between the first two sessions is shorter than between the last two sessions, while the opposite is true for a contracting schedule. To date, only few studies have tested all three or at least two distribution schedules against each other on educationally relevant time scales. Tsai (1927) had participants practice word pairs within a period of 9 days in multiple sessions using an expanding, a contracting, or an equal learning schedule. Memory was assessed with free recall tests 3 and 7 days later. On both tests retention was best when the material had been studied with an expanding learning schedule. A more recent study by Cull (2000) compared massed, equal, and expanding learning schedules for learning of word pairs. Practice took place within a period of 6 days and involved study-only, test-only, or test-with-feedback as learning events, with final cued recall tests administered 3 days or 8 days later. Irrespective of learning event, an overall benefit of distributed practice compared to massed practice was revealed and both expanding and equal schedules improved memory equally well. Gerbier and Koenig (2012) required participants to learn word-pseudoword pairs across 7 days. In their first experiment, participants engaged in study-only trails during practice. They found an advantage of expanding learning schedule compared to equal or contracting learning schedule 2 days later. In a follow-up study, participants indicated whether they recognized a presented word pair during relearning. This time, both expanding and equal learning schedules improved cued recall performance compared to a contracting schedule. In a recent study by Küpper-Tetzel, Kapler, and Wiseheart (2014) participants practiced word pairs using a contracting, equal, or expanding learning schedule for a period of 7 days with a test-with-feedback procedure. Assessing final memory performance immediately or after a 1-, 7- or 35-day RI, their research demonstrates that the optimal learning schedule depended on the length of the RI: On the immediate free recall test, all three learning schedules fared equally well. For the 1- or 7-day RI conditions, the contracting schedule outperformed the equal and expanding schedules, whereas in the longest RI condition, participants benefited more from an equal and expanding schedule than from a contracting one. Building on this finding, additional studies are needed that use long RIs and particularly look at benefits of different learning schedules for learning of educationally relevant material.

Exclusively focusing of equal learning schedules, another line of research investigated the influence of longer or shorter ISIs between learning sessions on long-term retention of verbal material (Bahrick, 1979; Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelps, 1987). In these studies, participants learned English-foreign language word pairs with a test-with-feedback procedure during multiple learning sessions (≥ 6 relearning sessions). Learning sessions were separated by ISIs varying between 1 day and 56 days. Word pairs were studied to criterion in each learning session and participants were tested after RIs ranging between 1 month and 8 years. As a result, people were found to learn faster when the material was restudied after shorter ISIs, whereas long-term retention improved considerably when study sessions were separated by longer ISIs. Bahrick et al. (1993) complement this finding by showing that fewer relearning sessions separated by longer ISIs increased memory performance to the same extent as more study sessions separated by shorter ISIs. It should be noted that all three studies confound the number of repetitions during relearning sessions and the ISI between sessions because participants studied the material during relearning sessions until reaching criterion of perfect performance. Due to forgetting, the number of needed repetitions to reach perfect performance must increase as ISI increases. Statistically controlling for the number of repetitions, Bahrick and Hall (2005) verified that the distribution of learning had an independent and positive effect on final retention. Moreover, they emphasized the importance of more flexible research designs to account for naturalistic settings where learners self-regulate their learning, e.g., by increasing study time or exposure to items. Taking up on this, participants in Kornell's (2009) study were allowed to allocate as much study time as they wanted to learn synonym word pairs with a computerized flashcard simulation. Again, practicing word pairs in a spaced rather than a massed fashion turned out to be advantageous, despite no difference in study times between conditions. Additionally, for a single learning session, Kornell advises against dividing a large stack of flashcards into many smaller ones because large flashcard stacks increase within-list item distribution, which promotes memory performance.

Beneficial effects of equally distributed practice were also found in applied studies. For example, Seabrook, Brown, and Solity (2005) and Ambridge, Theakston, Lieven, and Tomasello (2006) demonstrated benefits of distributed practice for acquisition of reading skills and complex sentence construction in 4- to 5-year-olds. In line with laboratory experiments, Bloom and Shuell (1981) presented evidence for the spacing effect and RI interaction. High school students took massed (30-minute unit/single day) or distributed (10-minute units/3 consecutive days) practice tests on vocabulary words during French class. On an immediate test, students in the two conditions did not differ in their memory performance, but on a test 4 days later, students in the spaced condition outperformed peers in the massed condition.

To sum up, laboratory as well as field studies provide converging evidence that the distribution of learning sessions is a promising method to maintain verbal information for a considerable period of time. Importantly, the positive effects on verbal learning were robust across different age groups, including children, high school students, young adults, and older adults, too (e.g., Simone, Bell, & Cepeda, 2012).

Distributed Practice Effect in Conceptual Learning in Mathematics and Science

Plenty of empirical evidence attests to the beneficial effects of distributed practice for rote learning of verbal material. Indeed, Roediger and Pyc (2012) argue that rote learning is important because it forms the basis on which to perform conceptual high-level operations. However, can distributed practice help conceptual learning, too?

In mathematics or science, the ability to understand how different pieces of information are interconnected and how knowledge can be applied to new problems are key competences. A nontrivial question is if retention of conceptual knowledge may profit from distributed practice similar to rote learning. In fact, once conceptual knowledge is acquired and applied successfully during initial learning, distributed practice may not produce any additional benefits. Assuming that the learner's comprehension of complex concepts is facilitated during an *extended* initial learning unit, it is possible that for conceptual learning massed practice is indeed superior to distributed practice.

In a study by Rohrer and Taylor (2006), participants were taught how to determine the number of permutations of letter sequences (e.g., *aaabbb*, Answer: *20*). Afterwards, they practiced five (spaced practice) or ten (massed practice) permutation problems during initial study session. The spaced practice group returned 1 week later to practice the other five problems. One or four weeks later, participants calculated permutations of new problems. The first finding was that participants in both conditions understood the mathematical concept and were able to successfully solve practice problems during initial learning. Second, participants in the spaced condition performed worse on the second half of the permutation problems than participants in the massed condition, suggesting that mathematical knowledge acquired 1 week earlier underwent forgetting. Interestingly though, after a 4-week RI, participants in the massed condition performed worse than those in the spaced condition.

Grote (1995) examined the effect of distributed practice for physics learning in high school. Practice was either massed into a single day or distributed across 20 consecutive days, and memory performance was assessed 2, 4, and 6 weeks later. On all tests students showed better retention for material that has been practiced in a distributed way. Moreover, Yazdani and Zebrowski (2006) demonstrated that homework assignments in high-school geometry class covering a percentage of previously-taught material were more effective than homework that contained only recently-taught material. An early study by Cook (1934) comes to the same conclusion as skill acquisition on a problem solving task was faster with massed practice than with distributed practice, but performance on the final tests were superior for the latter. Finally, Vlach and Sandhofer (2012) taught science concepts to 6-year-old elementary school children in four lessons occurring either on a single day (massed condition), two successive days (two lessons/day), or four successive days (one lesson/day). Their results confirmed that children who practiced on four successive days performed better on simple generalization questions, but also – and this is remarkable – on complex generalization questions tapping on the deep comprehension of the underlying abstract structure of concepts.

In conclusion, the few studies on conceptual learning in mathematics and science suggest that distributed practice has positive effects on retention of simple and complex knowledge as well as on mastering problem solving tasks. Since all studies to date have looked at spacing effects only, more studies are needed to enrich the data basis, especially those examining lag effects.

Theories of the Distributed Practice Effect

How can the distributed practice effect be explained? What are the mechanisms underlying this promising phenomenon often appraised by researchers? Can existing theories help us understand this strong learning effect? A theory of the distributed practice effect should be able to explain – and at best predict – the following findings: First, the benefit of spaced practice compared to massed practice on delayed recall tests. Second, the superiority of massed practice to spaced practice on immediate recall tests following practice. Third, the systematic interaction between ISI and RI found for rote learning in verbal recall tasks.

In the following, I evaluate theoretical approaches that have been suggested as explanations for the distributed practice effect: the study-phase retrieval hypothesis (Thios & D‘Agostino, 1976) and the contextual variability theory (Glenberg, 1979). In addition, I present the two-factor model (Verkoeijen, Rikers, & Schmidt, 2004), which combines both theories. In fact, these theories are often mentioned in research papers on the distributed practice effect, but seldom tested directly. This is due to the fact that, in general, proposed theories are quite weak in the sense that *strong* predictions are difficult to derive. To foreshadow, all of them offer plausible explanations for the advantage of spaced compared to massed practice. However, the systematic relationship between ISI and RI is poses a problem for most theories, particularly when it comes to offer precise predictions on when *exactly* the learner should reengage in practicing previously studied material.

Finally, I briefly outline the Multiscale Context Model (MCM; Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009) as an example for a computational memory model of the distributed practice effect and conclude with a new way to test theories of this effect using multinomial processing tree (MPT) models (Erdfelder et al., 2009).

Study-phase retrieval hypothesis

The study-phase retrieval hypothesis (Thios & D‘Agostino, 1976) is based on the observation that the distributed practice effect depends on the successful recognition of an item as repetition during its second presentation at practice (Bellezza & Young, 1989). It rests on the assumption that final memory performance benefits from distributed practice if during repeated studying of an item its previous learning episode is retrieved from memory (i.e., study-phase retrieval). This study-phase retrieval is assumed to be cued automatically by the present learning episode and to strengthen the memory trace (Benjamin & Tullis, 2010; Braun & Rubin, 1998). The positive and enhancing effect of retrieval from memory is, for instance, potentiated in the *testing effect*, where it has been shown that recalling during practice reduces forgetting of the to-be-learned material compared to restudying it (e.g., Carrier & Pashler, 1992; Roediger & Karpicke, 2006; Carpenter, Pashler, & Cepeda, 2009). Wahlheim, Maddox, and Jacoby (2013), for example, demonstrated that when study-phase retrieval is encouraged through instructions cued recall performance increases as a consequence. Study-phase retrieval is thought to be most effective when processing is maximally effortful, an idea originated by Hintzman’s (1974) deficient processing theory. In case of the distributed practice effect, this is true when the ISI is long. For example, Magliero (1983) showed in a within-list paradigm that longer ISIs between repetitions led to increased pupil dilations, a proxy for processing effort. At best, material should be reviewed right before it is forgotten, thereby ensuring successful, but maximally effortful processing. In line with this, Pyc and Rawson (2009) found that successful and more effortful retrievals during practice (longer response times indicating more effortful retrieval) – as a consequence of longer ISIs compared to shorter ISIs – improved memory performance on the final test more than successful, but less effortful retrievals.

However, study-phase retrieval can fail altogether if ISIs become too long or the context is changed and the material is modified between learning episodes (e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Durgunoglu, Mir, & Arino-Marti, 1993). This will harm memory performance on the final test. Consequently, the optimal time for relearning will depend on many factors, e.g., the study material, how well the material was encoded in the first place, variations in context between sessions, and learner characteristics.

In principle, the study-phase retrieval theory can explain that the memory performance function reaches a peak at a specific ISI and declines for longer ISIs due to forgetting between learning sessions. However, it has trouble accommodating the finding that the optimal ISI depends on the length of the RI. The study-phase retrieval hypothesis would suggest one optimal ISI, which depends on the abovementioned factors, but not on the RI. In addition, a prediction of the exact timing of when material should be optimally reviewed is hard to derive.

Contextual variability theory

Glenberg (1979) provided the most comprehensive version of the contextual variability theory – also referred to as encoding variability theory. The contextual variability theory proposes that a piece of information is stored in a memory trace along with *contextual components*. Contextual components can, for instance, be the physical context (e.g., features of the study setting, room temperature, smell, or noise), the time (e.g., early in the morning or late at night), a learner's inner state (e.g., positive or negative state), but also interitem relations (e.g., associations built between the material that is studied), or more subtle features of the learning task. As time goes by, contextual components are assumed to fluctuate (Estes, 1955).

Consequently, if a piece of information is repeatedly studied, a variety of different contextual components will be stored with its memory trace with the passage of time. The crux of this theory is that final memory performance will depend on the overlap between the contextual components that are present during the final test session and the ones stored in memory. In other words, contextual elements at test are assumed to function as a cue to retrieve the correct memory trace – and this cueing works best if contextual components match. The strongest asset of the contextual variability theory is that it predicts the optimal time to review material to depend heavily on the length of the RI. Imagine the simplest distributed practice design: Two learning sessions separated by an ISI and one test session after a specific RI. When the RI is short, the learner will benefit more from a short ISI than from a long ISI. That is because contextual components at test will have a greater overlap with the contextual components stored in memory during *both* learning sessions. However, when the ISI is long, contextual components at test will overlap to a great deal with the second learning session context, but not the first learning session context leading to attenuated performance. When the RI is long, contextual components at test will be at random. In this case, the match will be better when the memory trace contains maximally distinct contextual features, which is more likely for long ISIs compared to short ISIs. Hence, the contextual variability theory does an excellent job in explaining why massed practice is superior for immediate recall tests, while distributed practice is better for delayed recall tests. Also, the increase in optimal ISI with RI is compatible with this theory.

However, several studies have challenged the core assumption of the contextual variability theory: namely, that greater independence between learning events resulting in increased variation of a memory trace leads to better performance on a later test. Studies that deliberately manipulated variability at repetition often showed no increase in memory, and sometimes even a decrease in final test performance (e.g., Dempster, 1987; Maki & Hasher, 1975; Postman & Knecht, 1983). Thus, enhanced variability leading to more retrieval routes has to be relaxed as the *sole* mechanism for the distributed practice effect. The variety and type of possible contextual components that may or may not be stored in a memory trace or be present at test make it difficult to predict matching, and thereby retrieval likelihood.

Combining study-phase retrieval and contextual variability: The two-factor model

The two-factor model proposes that both contextual variability and study-phase retrieval mechanisms are responsible for the distributed practice effect. Young and Bellezza (1982) demonstrated that encoding variability did not enhance memory performance. Quite the contrary, memory performance declined when variability during encoding was introduced. They concluded that retrieval at final test is enhanced only when the initially formed memory trace is retrieved during repetition (study-phase retrieval) and enriched with contextual features that are sampled from the current study episode (contextual variability). Consequently, study-phase retrieval may fail if the current learning episode is too dissimilar to previous learning episodes. Verkoeijen et al. (2004; 2005) specified this theory further by suggesting that, at first, memory performance is dominated by the contextual variability component (i.e., increasing the ISI increases the variety of stored contextual information). This, in turn, increases memory performance on the final test due to better matching of contextual components. But the integration of new contextual information only occurs as long as study-phase retrieval from memory is successful. Thus, while the probability of contextual variability increases with ISI, the probability of successful study-phase retrieval decreases. If study-phase retrieval fails, no integration of new contextual information takes place harming final memory performance. Thus, these opposing mechanisms inevitably produce an inverted U-shaped memory function with increasing ISI.

Studies provided converging evidence for the two-factor model by demonstrating that moderately long ISIs are more beneficial than long ISIs (see Toppino, Hara, & Hackman, 2002; Verkoeijen et al., 2008) or by showing that a manipulation of contextual features on top of distributed practice harms memory performance (Verkoeijen et al., 2004). Since the two-factor model is, in principle, the combination of the previously presented theories, similar drawbacks apply. In particular, the optimal lag's dependency on the study-phase retrieval mechanism implies that this theory, too, cannot explain ISI-RI interactions.

Computational memory models

Computational memory models have been used to obtain a better understanding of the distributed practice effect and to predict the best time to reengage in learning of the to-be-learned material by appropriately specifying model parameters. I describe the recently developed Multiscale Context Model (MCM; Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009) as an example.

MCM is a combination of the Search of Associative Memory model (SAM; Raaijmakers, 2003) and the Predictive Utility Theory (Staddon, Chelaru, & Higa, 2002). SAM is essentially a formalization of the two-factor model. To revisit, retrieval of the previously stored memory trace is crucial for the integration of new context information and matching of contextual information is essential for successful retrieval at test. If study-phase retrieval fails, a new memory trace is formed instead, leading to an attenuation of the distributed practice effect. The Predictive Utility Theory is borrowed from habituation studies in animals and states that the time that elapses between repetitions dictates for how long the information will be maintained for the future. If a piece of information is reencountered shortly (a long time) after the previous occurrence, memory will store that piece of information in a way that it is maintained for shorter (longer) periods of time. Mozer et al. (2009) and Lindsey, Mozer, Cepeda, and Pashler (2009) show in their simulation studies that MCM can nicely fit empirical distributed practice data *post hoc*. However, a more recent study by Lindsey, Shroyer, Pashler, and Mozer (2014) presented a better model that integrates parameters for learner abilities, item difficulty, past study history, and forgetting. They demonstrate that when practice of foreign language vocabulary was distributed according to the assumptions of this new model, students showed superior memory performance on a test 1 month later compared to alternative models. The advantage of computational memory models is that – once their validity is established – they can be fed in computer programs and be used to determine optimal time to review the to-be-learned material (see the *Colorado Optimized Language Tutor* by Lindsey et al., 2014, for an example).

Taken together, MCM suggests that a certain degree of forgetting before material is reviewed has positive effects on long-term retention and that forgetting can depend on different factors that should be taken into consideration.

Testing distributed practice effect theories with Multinomial Processing Tree Models

Memory performance that is assessed during the final test session is always a product of different cognitive processes, e.g., encoding processes, maintenance processes to final test, as well as retrieval processes at test. However, these different processes may not be equally contributing to memory performance in the distributed practice paradigm. It might be the case that the advantage of distributed practice is particularly driven by one or a combination of two processes. Knowing the relative contributions of different memory processes allows drawing conclusions for the validity of theories because different theories emphasize different processes. For example, the study-phase retrieval hypothesis suggests that processes during *encoding* (i.e., retrieval of previously stored memory trace *during practice* leads to strengthening of that trace) play a particularly crucial role for the distributed practice effect. In contrast, the decisive factor for the contextual variability theory is the matching of contextual components *during testing* with the ones stored in memory, which highlights the role of *retrieval* processes during final test. MCM suggests enhanced *maintenance processes to the time of testing* as an important aspect for the distributed practice effect due to its Predictive Utility assumption.

In the past, multinomial processing tree (MPT) models (see Erdfelder et al., 2009), have been used to decompose memory performance into contributions of the underlying cognitive processes (e.g., bizarreness effect: Riefer & Rouder, 1992; recognition failure effect: Riefer & Batchelder, 1995). A recent study by Küpper-Tetzel and Erdfelder (2012) has applied MPT modeling to distributed practice data and examined the role of encoding, maintenance, and retrieval processes contributing to the shift in optimal ISI with increasing RI. They had participants engage in study-test-trials of word pairs in two learning sessions that were separated by a 0-, 1-, or 11-day ISI. Final free and cued recall tests were administered after RIs of 7 or 35 days. Küpper-Tetzel and Erdfelder found that memory performance after a 7-day RI followed an inverted U-shaped trend with a peak at a 1-day ISI. In contrast, memory increased linearly between the 0- and 11-day ISI conditions for the 35-day RI. Using their Encoding-Maintenance-Retrieval MPT model they revealed that in both RI conditions the probability of *encoding* increased between the 0- and 1-day ISI, but decreased between the 1- and 11-day ISI. This affected memory performance assessed 7 days later, but not 35 days later. The beneficial effects of a 11-day ISI for the 35-day RI condition was due to enhanced *maintenance processes to the time of testing* since probability of maintenance increased between the 0-day and the 11-day ISI condition. The probability of *retrieval* increased between the massed condition and the two distributed conditions, but did not differ between the 1- and 11-day ISI conditions. Thus, encoding processes largely affected performance for a 7-day RI, whereas maintenance processes were key for a 35-day RI. Consequently, a valid theory should consider the interplay of encoding and maintenance processes for explaining the systematic interaction between ISI and RI. Potential candidates that propose valid *processes* are the study-phase retrieval theory or MCM, while the contextual variability theory with its emphasis of retrieval processes at test is difficult to reconcile with the MPT-based findings.

Conclusion

This review provides an overview of empirical findings of the distributed practice effect that are highly relevant for educational practice. The superiority of practice distribution has been demonstrated for rote learning in verbal tasks, but also to some extent for conceptual learning in mathematics and science. In addition, all age groups seem to benefit from this easy-to-apply learning strategy as benefits have been revealed in young children, young adults as well as older adults. The most recent meta-analysis on this learning phenomenon attests to large effect sizes throughout (Cohen's $d \geq 1$, Cepeda et al., 2006). Thus, although most of the evidence comes from studies conducted in highly controlled laboratory settings, researchers are optimistic that the benefits of distributed practice are scalable to real-world settings, too. The few existing field studies confirm this hypothesis (e.g., Carpenter et al., 2009; Küpper-Tetzel et al., 2013; Sobel et al., 2011).

Still, more studies are needed to smooth the connection between research and practice. Besides field studies in educational settings, studies should use more authentic materials that require learners to obtain a deep understanding of the to-be-studied subject. Also, future experiments should take learners' study habits and aspects of self-regulated learning more into account. Keeping study time and the number of presentations constant for all participants is important for the experimental design, but limits our understanding of learning in the real world and ignores the important momentum of the learner's decision (see Bahrick & Hall, 2005; Kornell, 2009, for a similar argument). A learner's belief about her own learning performance influences subsequent learning decisions, which may affect memory performance. Cohen, Yan, Halamish, and Bjork (2013), Kornell and Metcalfe (2006), and Son (2010), for example, have investigated the circumstances under which people decide to space or mass practice of an item. Shifting control to the learner certainly represents a better approach to an authentic learning environment and opens new research opportunities on how, e.g., students' learning decisions affect memory performance. Developing a new research program along these lines promises to improve the validity of recommendations from research and to enhance the confidence in implementing the distributed practice effect in education.

Although these points are important and should be emphasized in future research, I argue that another crucial aspect is to obtain a better understanding of the underlying mechanisms of the distributed practice effect. This review posits that existing theories cannot convincingly account for the different empirical findings, and predictions about the optimal point in time to review material are hard to derive. One option could be to rely on computational memory models that are appropriately parameterized to optimize scheduling (Lindsey et al., 2013). Another option could be to approach the distributed practice effect from a process-oriented perspective and to examine the underlying cognitive processes of the distributed practice effect using MPT models (Küpper-Tetzel & Erdfelder, 2012) to draw conclusions about the processes that a valid theory should embrace. This can confirm existing theories, or at least elements thereof, but can also motivate the development of new theories.

Although we are dealing with a very old learning phenomenon we have not yet obtained a full understanding of the distributed practice effect. Implementing the distributed practice effect as a learning strategy more comprehensively in real-world education will probably only take place once our understanding of this effect improves. If, for instance, forgetting between learning sessions were the crucial process for beneficial distributed practice effects, instructors could plan review sessions of different material accordingly by, e.g. revisiting material that is forgotten faster after a shorter ISI. Thus, understanding the distributed practice effect may increase not only the probability, but more importantly the quality of implementation. In a recently published paper in *Science* Koedinger, Booth, and Klahr (2013) stated that "theoretical work can offer insight into when an instructional choice is dependent on a learning [objective]" (p. 936). Without a consistent theoretical framework, providing advice beyond the exact experimental context is difficult as results are tied to the specific designs used to produce the empirical results. For applied settings, which are characterized by a variety of factors that are often hard to predict, empirical results on the distributed practice effect alone without a clear theoretical underpinning can hardly ever suffice to provide flexible hands-on advice for instructors and learners. Thus, future research on the distributed practice effect should focus on validating existing theories as well as on developing new ones to put this strong effect on strong theoretical grounds.

Acknowledgments

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). I thank Patrick Bayer and Michael Cahill for helpful comments on an earlier version of this article.

References

Ambridge, B., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2006). The distributed learning effect for children‘s acquisition of an abstract syntactic construction. *Cognitive Development*, *21*, 174–193. doi: 10.1016/j.cogdev.2005.09.003

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. Journal of Consumer Research, 32, 266–276. doi: 10.1086/432236

Austin, S. D. M. (1921). A study in logical memory. *The American Journal of Psychology*, *32*, 370–403. doi: 10.2307/1414001

Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools*, *3*, 195–198. doi: 10.1002/1520-6807(196607)3:3<195::AID-PITS2310030302>3.0.CO;2-X

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*, 296–308. doi: 10.1037/0096-3445.108.3.296

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316–321. doi: 10.1111/j.1467-9280.1993.tb00571.x

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, *52*, 566–577. doi: 10.1016/j.jml.2005.01.012

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344–349. doi: 10.1037/0278-7393.13.2.344

Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 990–997. doi: 10.1037/0278-7393.15.5.990

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228–247. doi: 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*, 635–650. doi: 10.1017/S0142716410000172

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

Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence from once-presented words. *Memory*, *6*, 37–65. doi: 10.1080/741941599

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. doi: 10.1002/acp.1507

Carpenter, S., Cepeda, N., Rohrer, D., Kang, S. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*, 369-378. doi: 10.1007/s10648-012-9205-z

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. Memory & Cognition, 20, 633–642. doi: 10.3758/BF03202713

Castillo, M. (2013, January 10). What's the best way to learn? Psychologists tackle studying techniques. *CBS News*. Retrieved from www.cbsnews.com.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*, 236–246. doi: 10.1027/1618-3169.56.4.236

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380. doi: 10.1037/0033-2909.132.3.354

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*, *19*, 1095–1102. doi: 10.1111/j.1467-9280.2008.02209.x

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*, 1682-1696. doi: 10.1037/a0032425

Cook, T. W. (1934). Massed and distributed practice in puzzle solving. *Psychological Review*, *41*, 330–355. doi: 10.1037/h0074350

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235. doi: 10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

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. In Brian H. Ross (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 63–147). London: Academic Press.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162–170. doi: 10.1037/0022-0663.79.2.162

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*, 627–634. doi: 10.1037/0003-066X.43.8.627

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*, 4–58. doi: 10.1177/1529100612453266

Durgunoglu, A. Y., Mir, M., & Arino-Marti, S. (1993). Effects of repeated readings on bilingual and monolingual memory for text. *Contemporary Educational Psychology*, *18*, 294–317. doi: 10.1006/ceps.1993.1022

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. Oxford England: Dover.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial Processing Tree Models. *Zeitschrift für Psychologie / Journal of Psychology*, *217*, 108–124. doi: 10.1027/0044-3409.217.3.108

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369–377. doi: 10.1037/h0046888

Gay, L. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology*, *64*, 171–182. doi: 10.1037/h0034595

Gerbier, E., & Koenig, O. (2012). Influence of multiple-day temporal distribution of repetitions on memory: A comparison of uniform, expanding, and contracting schedules. *The Quarterly Journal of Experimental Psychology*, 514-525. doi: 10.1080/17470218.2011.600806

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112. doi: 10.3758/BF03197590

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition*, *8*, 528–538. doi: 10.3758/BF03213772

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology*, *79*, 198–199. doi: 10.1037/0022-0663.79.2.198

Gordon, K. (1925). Class results with spaced and unspaced memorizing. *Journal of Experimental Psychology: General*, *8*, 337–343. doi: 10.1037/h0064623

Grote, M. G. (1995). Distributed versus massed practice in high school physics. *School Science and Mathematics*, *95*, 97–101. doi: 10.1111/j.1949-8594.1995.tb15736.x

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

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

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. Oxford, England: Lawrence Erlbaum.

Koedinger, K. R., Booth, J. L. , & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*, 935-937. doi: 10.1126/science.1238056

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*, 1297–1317. doi: 10.1002/acp.1537

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609–622. doi: 10.1037/0278-7393.32.3.609

Krug, D., Davis, T., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology*, *82*, 366–371. doi: 10.1037/0022-0663.82.2.366

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, *20*, 37–47. doi: 10.1080/09658211.2011.631550

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2013). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*, 1-16. doi: 10.1007/s11251-013-9285-2

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Expanding, contracting, and equal learning schedules? The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition.* Advance online publication. doi: 10.3758/s13421-014-0394-1

Lawton, T. W., Cronin, J. B., & Lindsell, R. P. (2006). Effect of interrepetition rest intervals on weight training repetition power output. *The Journal of Strength & Conditioning Research*, *20*, 172-176.

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological Science*. Advance online publication. doi: 10.1177/0956797613504302

Lindsey, R., Mozer, M. C., Cepeda, N. J., & Pashler, H. (2009). Optimizing memory retention with cognitive models. In A. Howes, D. Peebles, & R. Cooper (Eds.), *Proceedings of the Ninth International Conference on Cognitive Modeling* (ICCM 2009). Manchester, UK: ICCM.

Magliero, A. (1983). Pupil dilations following pairs of identical and related to-be-remembered words. *Memory & Cognition, 11*, 609-615. doi: 10.3758/BF03198285

Maki, R. H., & Hasher, L. (1975). Encoding variability: A role in immediate and long-term memory? *The American Journal of Psychology, 88*, 217–231.

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

Moulton, C.-A. E., Dubrowski, A., Macrae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Annals of surgery*, *244*, 400–409. doi: 10.1097/01.sla.0000234808.85789.6a

Mozer, M. C., Pashler, H., Cepeda, N. J., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems* (pp. 1321–1329). La Jolla, CA: NIPS Foundation.

Paul, A. M. (2011, September 10). The trouble with homework. *The New York Times*. Retrieved from http://www.nytimes.com.

Postman, L., & Knecht, K. (1983). Encoding variability and retention. Journal of Verbal Learning and Verbal Behavior, 22, 133–152. doi: 10.1016/S0022-5371(83)90101-9

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437–447. doi: 10.1016/j.jml.2009.01.004

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science: A Multidisciplinary Journal*, *27*, 431–452. doi: 10.1207/s15516709cog2703_5

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*, 70–80. doi: 10.1037/0022-0663.97.1.70

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. doi: 10.1037/h0040734

Riefer, D. M., & Batchelder, W. H. (1995). A multinomial modeling analysis of the recognition-failure paradigm. *Memory & Cognition*, *23*, 611–630. doi: 10.3758/BF03197263

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*, 601–611. doi: 10.3758/BF03202710

Roediger, H. L., & 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*, 242–248. doi: 10.1016/j.jarmac.2012.09.002

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

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

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498. doi: 10.1007/s11251-007-9015-8

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

Shea, C. H., Lai, Q., Black, C., & Park, J.-H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science*, *19*, 737–760. doi: 10.1016/S0167-9457(00)00021-X

Simmons, A. L. (2012). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education*, *59*, 357–368. doi: 10.1177/0022429411424798

Simone, P. M., Bell, M. C., & Cepeda, N. J. (2012). Diminished but not forgotten: Effects of aging on magnitude of spacing effect benefits. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68,* 674-680. doi: 10.1093/geronb/gbs096

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

Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 255–262. doi: 10.1037/a0017892

Staddon, J. E. R., Chelaru, I. M., & Higa, J. J. (2002). Habituation, memory and the brain: The dynamics of interval timing. *Behavioural Processes*, *57*, 71–88. doi: 10.1016/S0376-6357(02)00006-2

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837–848. doi: 10.1002/acp.1598

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning & Verbal Behavior*, *15*, 529–536. doi: 10.1016/0022-5371(76)90047-5

Toppino, T. C., Hara, Y., & Hackman, J. (2002). The spacing effect in the free recall of homogeneous lists: Present and accounted for. *Memory & Cognition*, *30*, 601–606. doi: 10.3758/BF03194961

Tsai, L. S. (1927). The relation of retention to the distribution of relearning. *Journal of Experimental Psychology: General*, *10*, 30–39. doi: 10.1037/h0071614

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*, 685–695. doi: 10.1002/acp.1388

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 796–800. doi: 10.1037/0278-7393.30.4.796

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology*, *52*, 257–263. doi: 10.1027/1618-3169.52.4.257

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*, 1137–1144. doi: 10.1111/j.1467-8624.2012.01781.x

Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 94-105. doi: 10.1037/a0034055

Yazdani, M. A., & Zebrowski, E. (2006). Spaced reinforcement: An effective approach to enhance the achievement in plane geometry. *Journal of Mathematical Sciences and Mathematics Education, 1*, 37-43.

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 545–559. doi: 10.1037/0278-7393.8.6.545