

Running head: CONTRACTING, EQUAL, AND EXPANDING LEARNING SCHEDULES

The final publication is available at
<http://link.springer.com/article/10.3758/s13421-014-0394-1>

Contracting, Equal, and Expanding Learning Schedules: The Optimal Distribution of Learning
Sessions Depends on Retention Interval

Carolina E. Küpper-Tetzel^{1,2}, Irina V. Kapler^{3,4}, and Melody Wiseheart^{3,4}

¹ Center for Integrative Research on Cognition, Learning, and Education (CIRCLE)
Washington University in St. Louis, USA

² Department of Psychology
University of Mannheim, Germany

³ Department of Psychology
York University, Canada

⁴ LaMarsh Centre for Child and Youth Research
York University, Canada

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
Unites States
Phone: +1- 314 935-6142
Email: carolina.kuepper-tetzel@wustl.edu

Abstract

In laboratory and applied learning experiments, researchers have extensively investigated the optimal distribution of two learning sessions (i.e., initial learning and one relearning session) for the learning of verbatim material. However, research has not yet provided a satisfying and conclusive answer to the optimal scheduling of three learning sessions (i.e., initial learning and two relearning sessions) across educationally relevant time intervals. Should the to-be-learned material be repeated at decreasing intervals (contracting schedule), constant intervals (equal schedule), or increasing intervals (expanding schedule) between learning sessions? Different theories and memory models (e.g., study-phase retrieval theory, contextual variability theory, ACT-R, and the Multiscale Context Model) make distinct predictions about the optimal learning schedule. We discuss the extant theories and derive clear predictions from each of them. To test these predictions empirically, we conducted an experiment in which participants studied and restudied paired associates with a contracting, equal, or expanding learning schedule. Memory performance was assessed immediately, 1 day, 7 days, or 35 days later with free and cued recall tests. Our results reveal that the optimal learning schedule is conditional on the length of the retention interval: A contracting learning schedule was beneficial for retention intervals up to 7 days, but both equal and expanding learning schedules were better for a long retention interval of 35 days. Our findings can be accommodated best by the contextual variability theory and indicate that revisions are needed to existing memory models. Our results are practically relevant and implications for real-world learning are discussed.

Contracting, Equal, and Expanding Learning Schedules: The Optimal Distribution of Learning Sessions Depends on Retention Interval

Students face a serious problem when it comes to the maintenance of once-learned material: They quickly forget a large portion of the knowledge that they have acquired at school and cannot access it later on when they need it (Bahrick & Hall, 1991). Consequently, interest in learning strategies that promise long-lasting knowledge maintenance has been immense. Researchers in cognitive psychology have revealed approaches that optimize learning, enhance memory performance, and reduce forgetting (see Pashler, Rohrer, Cepeda, & Carpenter, 2007, for a comprehensive review). One such learning strategy is the distributed practice effect — a learning phenomenon that dates back to research conducted by Hermann Ebbinghaus in 1885 (Ebbinghaus, 1885/1964). It refers to the finding that final memory performance is improved if learning sessions are distributed in time rather than being massed into a single study episode (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

The simplest research design to investigate the distributed practice effect consists of two learning sessions (an initial learning session, in which the material is studied for the first time, and a relearning session, in which material is revisited) and a final test session (in which memory performance of the to-be-learned material is assessed). The time interval between learning sessions is referred to as the interstudy interval (ISI), and the time interval between the last learning session and the final test session is referred to as the retention interval (RI). The optimal distribution of an initial learning session and one relearning session (i.e., of two learning sessions in total) has been intensively examined for the learning of verbatim material in the laboratory as well as in applied classroom-based studies (e.g., Cepeda et al., 2009; Glenberg & Lehmann, 1980; Küpper-Tetzel & Erdfelder, 2012; Küpper-Tetzel, Erdfelder, & Dickhäuser, 2013).

Those studies revealed that the optimal time to review material depends heavily on the length of the RI. More precisely, verbal material has a higher probability of being maintained for a long time in memory (e.g., an RI of 1 month or longer) if the ISI between initial learning and a review session is long, too (e.g., 11 days or longer). However, if the RI is short (e.g., 1 week), memory will benefit more from a short ISI (e.g., 1 day). Thus, the optimal ISI for relearning verbal material increases as the RI increases. Although this finding already has potential practical implications (Küpper-Tetzl et al., 2013), a more pressing issue concerns the question of how more than two learning sessions should be distributed optimally across educationally relevant time intervals, such as several days or weeks, in order to enhance long-term memory. Understanding the effects of two or more relearning episodes is particularly relevant in real-world settings, where learners are likely to revisit study material more than once in order to improve maintenance of the topic.

In general, multiple relearning sessions allow for the investigation of three distinct classes of learning schedule: contracting, equal, and expanding. In a contracting learning schedule the ISI between learning sessions decreases across time; in an equal learning schedule the ISI between learning sessions is constant; and in an expanding learning schedule the ISI between learning sessions increases across time. The three learning schedules are shown in Figure 1.

Empirical evidence for the effect of learning schedules

Most experiments that have compared different learning schedules have manipulated them within one experimental session. That is, the ISIs between relearning opportunities were quite short and filled with practice of intervening material or other tasks. This design will be referred to as a *within-session design*. These studies have produced equivocal findings in regard

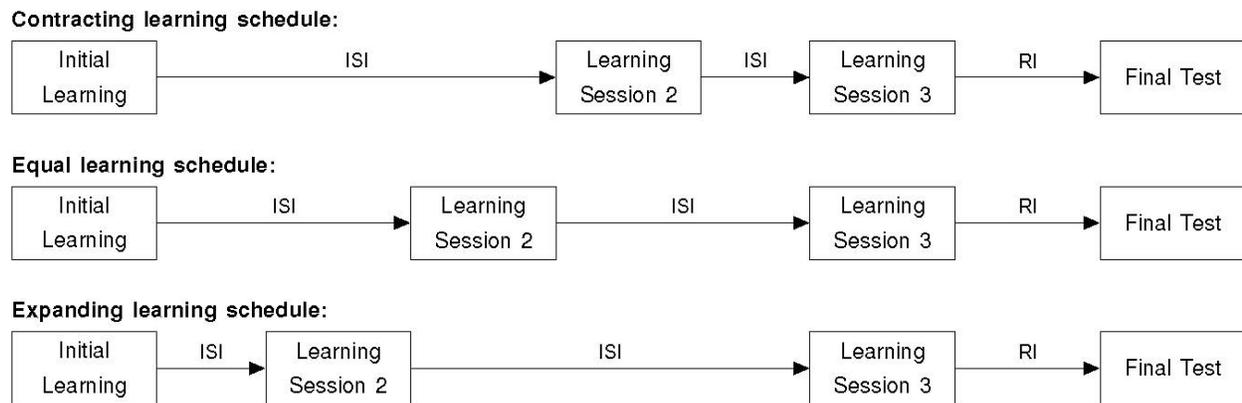


Figure 1. Contracting, equal, and expanding learning schedules, with time on the x-axis (ISI, interstudy interval). The final test session occurs after a fixed retention interval (RI) subsequent to Learning Session 3.

to the advantage of a specific learning schedule over the others (see Balota, Duchek, & Logan, 2007, for a review).

In the aftermath of the famous experiments by Landauer and Bjork (1978) – which report an advantage of expanding ISIs for when material is repeatedly tested without feedback and an advantage of equal ISIs for when material is repeatedly presented and read – research often has focused on pitting equal and expanding schedules against each other, ignoring the contracting condition (e.g., Carpenter & DeLosh, 2005; Cull, Shaughnessy, & Zechmeister, 1996; Karpicke & Roediger, 2010). Taken together, findings from these studies do not point to a general advantage of one learning schedule over the other. Two factors have been discussed as moderators for optimal learning schedule: learning event and RI. Cull et al. (1996), for example, showed that expanding intervals worked best for repeated tests without feedback, but found no difference between equal and expanding intervals when repeated tests with feedback were used. However, subsequent studies have refuted the moderating role of feedback on optimal learning schedule. For example, Cull (2000, Experiment 1) could not replicate the effect of feedback on optimal learning schedule. Similarly, Karpicke and Roediger (2010) found no evidence that the

advantage of any of the two schedules depended on whether learning tests were followed by feedback. In fact, in all of their experiments both equal and expanding intervals produced comparable outcomes. Carpenter and DeLosh (2005) compared repeated tests without feedback versus repeated studying at expanding and equal intervals and did not detect a reliable interaction between the two factors. Thus, literature pointing to learning event as a moderator is to date unclear and requires further investigation.

The role of RI in moderating the effectiveness of expanding versus equal learning schedules is more consistent. Karpicke and Roediger (2007), for instance, found an advantage of expanding schedules when the final test occurred after a short 10-min RI, but an advantage of equal learning schedules when the RI was 2 days. Logan and Balota (2008) confirmed this interaction between learning schedule and RI.

In a recent study by Karpicke and Bauernschmidt (2011), participants worked on study-test trials administered at contracting, equal, and expanding intervals in order to practice paired associates. In addition to the relative distribution of learning (i.e., contracting, equal, expanding) they also varied the length of the absolute learning interval. More precisely, the three different learning schedules were carried out either within a short learning interval (i.e., within 15 learning trials) or within a long learning interval (i.e., within 90 learning trials). On the final test occurring 1 week later, a clear effect of absolute learning interval emerged. That is, memory performance was better when learning was distributed across a long as compared to a short period of time. However, the specific learning schedule – whether it was equal, expanding, or contracting – did not affect memory performance.

The question that arises is whether these null effects of learning schedule would also hold true when learning sessions were separated by educationally relevant intervals, in the range of days or weeks. We refer to this as a *between-sessions design*. It is likely that the effect of

different learning schedules would emerge more clearly when learning episodes were distributed across days and not just interrupted by intervening material within a list, especially against the backdrop of theoretical explanations (outlined below) that emphasize forgetting processes between learning opportunities as a crucial aspect for the distributed practice effect (Lindsey, Mozer, Cepeda, & Pashler, 2009). Clearly, forgetting processes are more pronounced when the ISI between learning sessions is of meaningful length. Also, theoretical explanations such as inattention can only apply to single-session designs, because working memory by definition does not operate on a multiday scale, and Cepeda et al. (2006) demonstrate a lack of scale invariance through the change in optimal ISI to RI ratio as a function of scale (i.e., RI), whether time is treated linearly or logarithmically. Moreover, most existing studies have focused more on tests without feedback as learning event (Balota et al., 2007) and less on restudying or tests with immediate review of the correct answer. Karpicke, Butler, and Roediger (2009), however, argued that learners typically engage in rereading or testing with corrective feedback instead of self-testing without feedback, in line with evidence that feedback is necessary for error correction (Pashler, Cepeda, Wixted, & Rohrer, 2005). Consequently, it is important to investigate the effect of different learning schedules with ecologically valid learning techniques.

Only a few studies have investigated different learning schedules using a between-sessions design, in which restudying or tests with feedback were used as learning method (Cull, 2000; Gerbier & Koenig, 2012; Tsai, 1927). In an early study by Tsai (1927), participants restudied word pairs in multiple sessions using expanding, contracting, or equal learning schedules. Practice was distributed across 11 days. On a free-recall test administered after RIs of 3 and 7 days, participants performed best when the material had been studied with an expanding learning schedule. A study by Cull (2000, Experiments 2 and 3) compared the effects of massed, equal, and expanding learning schedules on paired-associate learning. During practice, word pairs

were either restudied, tested with feedback, or tested without feedback. Practice was spaced across a period of 6 days. Memory performance was assessed with a cued-recall test either 3 or 8 days after learning. Not surprisingly, the results revealed an overall benefit of distributed relative to massed practice, but both equal and expanding learning schedules improved memory to the same extent, and learning schedule did not interact with whether word pairs were studied or tested during practice. In a study by Gerbier and Koenig (2012), word-nonword pairs were studied with expanding, equal, or contracting intervals over a period of 7 days using study-only trials. Two days later, performance on a cued-recall test was better after participants had studied at expanding intervals. However, when participants' recognition of each word pair was tested during practice, expanding and equal schedules did not differ with regard to cued-recall performance measured 2 days later, and both schedules were superior to a contracting learning schedule.

Theoretical explanations for the effect of learning schedules

Empirical data paint a mixed picture and fail to offer a comprehensive conclusion as to which learning schedule is the most beneficial for remembering information over long time intervals. Is there a theoretical basis for predicting differences in memory performance as a consequence of changes in learning schedule distribution? In the following section, we highlight the two most discussed theories for the distributed practice effect – the study-phase retrieval hypothesis and contextual variability theory – and we present the Multiscale Context Model (MCM; Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009) as a computational model that integrates both approaches. The predictions from MCM will be contrasted against predictions from the Adaptive Character of Thought – Rational model (ACT-R; Pavlik & Anderson, 2008), which has

been repeatedly suggested as a potential explanation for the distributed practice effect. The present experiment aims at testing the validity of these two memory models.

The classical explanation for expanding learning schedule superiority has been that expanding intervals result in a higher probability of continuous retrieval success of the to-be-learned material during practice (Landauer & Bjork, 1978). Put in opposite terms, equal and contracting schedules are more likely to result in forgetting because memory traces are weak after initial learning and the memory strength boost (i.e., in form of a relearning session) needs to come sooner. Expanding intervals maintain a high memory performance level throughout the learning phase and, in turn, benefit retention on the final test. Whereas it is true that an expanding learning schedule leads to better memory performance during learning than does an equal learning schedule, this benefit has not always translated into improved long-term retention. Quite the contrary: Delaying the first relearning opportunity – as is the case in equal learning schedules – has led to improved long-term retention (Karpicke & Roediger, 2007; Logan & Balota, 2008).

The study-phase retrieval theory (Thios & D'Agostino, 1976) suggests that distributed practice is most beneficial if a repeated presentation of an item is successfully recognized as such during practice (Bellezza & Young, 1989; Braun & Rubin, 1998; Toppino, Hara, & Hackman, 2002), but, at the same time, its processing should be effortful. In principle, the second encounter with an item works as a cue for beneficial study-phase retrieval. When it is effortful, successful study-phase retrieval during learning strengthens the memory trace and translates into enhanced memory performance on the final test. However, if study-phase retrieval is too easy because it occurs after a short ISI, or too hard because the to-be-learned material has been forgotten during a long ISI, final retention will be attenuated. This hypothesis may explain why delaying the first restudying opportunity has beneficial effects on memory. In other words, in the framework of the study-phase retrieval theory, to-be-learned material should be reactivated right before it is

forgotten. Thus, depending on the forgetting rate of the to-be-learned material, this theory might favor equal, or even contracting, learning schedules (for well-learned and slowly decaying material) over expanding learning schedules. Study-phase retrieval theory suggests that there is one optimal ISI set for all RIs, which is whatever ISI set maximizes encoding strength during learning.

The contextual variability theory (Glenberg, 1979), a second major explanation for the distributed practice effect, states that an item is always stored along with its context (e.g., other items learned at nearly the same time, emotional state, visual environment, etc.). These contextual components undergo natural fluctuations over time (Estes, 1955). As a consequence, a greater number of different contextual components are likely to be stored in association with the memory trace of an item as ISI increases, because of random contextual drift and decreasing likelihood for any given contextual element to overlap in state as delay increases. The matching between the contextual components present at test and those stored in a memory trace determines the retrieval probability of an item at test. The contextual features at test serve as retrieval cues to access the memory trace, and the greater the match between encoding and test contextual states, the greater is the probability of successful retrieval. The contextual variability theory predicts that the optimal learning schedule depends on the length of the RI. The contextual components present at test after a long RI (e.g., 1 month) will consist of a random sample of components. Hence, retention should benefit most from an equal learning schedule, because the contextual components that were stored during learning are maximally distinct from each other, allowing for greater contextual variety in the memory trace (e.g., $X_a - - - X_b - - - X_c - - - - - Test_{random}$)¹. In contrast, when the RI is short (e.g., 1 day), the contextual components at test will probably be

¹ The X s indicate different learning sessions and $Test$ represents the final test session. Subscripts symbolize the contextual components stored in memory and present at the final test. Hyphens represent time.

more similar to the contextual components stored during the last learning session(s).

Consequently, memory retrieval should benefit most from a contracting learning schedule ($X_a - - - X_b X_b$ -Test_b), because the matching probability is higher when two learning sessions are both so close to the final test.

Mozer et al. (2009) developed MCM, which in essence represents a hybrid of the study-phase retrieval and contextual variability theories (see also Raaijmakers, 2003), with the addition of the *predictive utility assumption*, a concept derived from models of habituation in animals (Staddon, Chelaru, & Higa, 2002). More precisely, MCM assumes that the to-be-learned material is stored along with contextual information and that the integration of contextual features to existing memory traces during repeated practice is conditional upon successful retrieval of the respective traces (i.e., study-phase retrieval). The degree of matching between contextual features during final test and those stored in the memory trace will determine performance on the test (i.e., contextual variability theory). Most importantly, the predictive utility assumption states that the time that elapses before the reencounter of a piece of information determines how long it will be maintained in memory. If the ISI is short, the material will be stored in a way that ensures best maintenance for short RIs. In contrast, if the ISI is long, the material will be stored so that it is maintained for a longer period of time (i.e., for longer RIs).

ACT-R, a second mathematical instantiation of the distributed practice effect, proposes that each time a piece of information is practiced, a new memory trace is stored, and that its activation decays over time, following a power law function (Pavlik & Anderson, 2008). Most importantly for the distributed practice effect, memory trace activation of a currently studied item will decay faster when the sum of activation of previously stored memory traces for that item is still large. This feature occurs when the ISI between two learning sessions is short. Thus, in ACT-R long-term retention should benefit more from longer ISIs between learning sessions.

In a simulation study, Lindsey, Mozer, Cepeda, and Pashler (2009) parameterized and tested MCM against ACT-R (Pavlik & Anderson, 2008) for their predictions of memory performance in a situation with three learning sessions. ACT-R and MCM make opposing predictions about the optimal distribution of practice. Whereas ACT-R predicts contracting intervals to be best, irrespective of the RI, MCM favors equal and expanding intervals depending on the RI. Equal learning schedules were best when the RI was 2 hours and shorter; expanding learning schedules fared better for longer RIs (≥ 1 day).

Overview of the present experiment

The presented mathematical models call for an *experimentum crucis*. The present experiment aims to broaden the empirical basis concerning the question of the optimal learning schedule for verbatim learning and to test predictions from the abovementioned models. Contrary to previous studies that have used RIs of only up to 8 days, memory performance was also assessed after a longer RI of 35 days. In order to test for the dependency of the optimal learning schedule on the length of the RI, we used four different RI conditions – namely 0-, 1-, 7-, and 35-day RIs. In addition, participants engaged in tests with feedback during learning instead of tests without feedback. As we described before, most studies have focused on providing tests during practice without corrective feedback. We believe that tests with feedback represent a more ecologically valid learning method than does giving tests without any form of feedback (see also Karpicke et al., 2009), and that it results in higher final long-term recall rates (Karpicke & Roediger, 2010, Experiment 2), which is especially important when examining an RI of 35 days. Many studies have emphasized the comparison between equal and expanding learning schedules and dismissed the potential benefits of a contracting learning schedule. However, as the discussion of different theories has shown, contracting learning schedules may have merit.

Against this backdrop, it is crucial to test all three possible learning schedules that can be constructed with three learning sessions – namely, contracting, equal, and expanding schedules. Therefore, we implemented all three learning schedules in the present experiment and tested them against each other.

Method

Participants

A total of 243 participants began the experiment. Of these, 218 completed the entire experiment (i.e., came to all experimental sessions)². Eight participants were excluded from all analyses due to failure to comply with instructions, extreme outlier status, lack of motivation, or cheating during learning. The remaining 210 participants were on average 23 years old ($SD = 4$, range = 18-40), 64% were female, and 98% rated their English language proficiency as “native,” “very good,” or “good.” The majority of the participants were undergraduate (82%) or graduate students (11%) enrolled at York University. They received course credits or a payment of CAN\$30 for their participation.

Materials

The material consisted of 56 concrete and highly familiar nouns (word length ranged between three to six letters), which were combined to produce 28 word pairs holding no obvious semantic association to each other.

Design

We manipulated *Learning Schedule* and *RI* in a 3 (Learning Schedule: Expanding, Contracting, or Equal) x 4 (RI: 0, 1, 7, or 35 days) between-subjects design. Participants underwent one Initial Learning Session and two relearning sessions, which were distributed

² Of the 25 participants who did not complete all experimental sessions, six were in the expanding-35-day RI condition, five were in the equal-7-day RI condition, and four were in the expanding-0-day RI condition. The rest were evenly distributed (one or two per condition) across all other conditions, except for the expanding-7-RI condition, which had no dropouts.

across a period of 7 days. In the contracting condition, the second learning session occurred 5 days after initial learning, and the third learning session occurred 1 day after the second learning session (i.e., 5- and 1-day ISIs). In the equal condition, the second learning session took place 3 days after initial learning, and the third learning session occurred 3 days after the second learning session (i.e., 3- and 3-day ISIs). In the expanding condition, the second learning session occurred 1 day after initial learning, and the third learning session took place 5 days after the second learning session (i.e., 1- and 5-day ISIs). Revisit Figure 1 for a visualization of the different learning schedules. Fifteen minutes (0-day RI), 1 day, 7 days, or 35 days after the third learning session, participants attended the final test session. Participants were assigned randomly to one of the 12 experimental conditions. The shortest experimental condition took 7 days to complete (i.e., any learning schedule combined with a 0-day RI). The longest experimental condition lasted 42 days (i.e., any learning schedule combined with a 35-day RI). The number of participants per condition ranged between 15 and 20.

Procedure

The participants attended three learning sessions and one final test session. All of the experimental sessions were computer-based, and four participants could be run simultaneously at different computers.

Participants read and signed consent forms and started the experiment at a computer. The experiment began with a presentation of the 28 word pairs. Each word pair was presented for 5 s, separated by a 750-ms interstimulus interval. The presentation of word pairs was randomized for each participant. Afterward, participants worked for 2 min on an arithmetic task as a distractor activity. Then they studied each word pair to a criterion of two correct answers in a cued-recall-with-feedback procedure. Specifically, participants were presented with the left word of a pair (cue) and were asked to type the corresponding right word (target). This test was self-paced.

Immediately after confirming their answer to the cue word, participants received feedback (“That was correct!” or “That was incorrect!”) and the correct word pair was displayed for 5 s.

Participants were presented with cue words in a random order until they had provided the correct target word for each cue twice, with items dropping from the testing rotation as they were learned to criterion. Cue words were presented in a random order on each trial until no more cue words were left. After a 2-min arithmetic distractor task, the Initial Learning Session concluded with a free-recall test and then a cued-recall test. No feedback was provided for these last two tests. On the free-recall test, participants were asked to recall all of the word pairs from memory without the cue being provided. They were encouraged to write down all words that they could recall, even if they only remembered one word of a pair. They were allotted a maximum of 5 min for the free-recall test, but could advance in the experiment once they were done recalling items. Then a cued-recall test followed, on which participants were presented with each cue word and prompted to type the corresponding target word. This test was self-paced. Participants were reminded of their second learning session and dismissed.

Learning Session 2 and Learning Session 3 took place after the predetermined randomly chosen ISI according to the participants’ assignment to a contracting (5- and 1-day ISIs), equal (3- and 3-day ISIs), or expanding (1- and 5-day ISIs) condition. Both relearning sessions started with one trial of cued recall with feedback. After a 2-min unrelated arithmetic task, participants were tested on a free-recall test and then a cued-recall test (both without feedback) for their memory after relearning. Afterward, participants were reminded of their next appointment and dismissed.

After their respective RI, participants completed the final test session. Participants in the 0-day RI condition played Sudoku for 15 min after their third session and continued with the test session on the same day. During the test session participants completed a final free-recall test and

then a final cued-recall test. No feedback was provided, and both tests were self-paced in the way described before. At the end of the experiment, participants were compensated for their participation, debriefed, and if they desired, were signed up to receive an email with promising research-based learning strategies for their own use (e.g., ideas on how one could implement spacing and testing strategies into one's own study habits).

Results

Memory performance during learning

The participants studied all 28 word pairs during their Initial Learning Session until reaching a criterion of two correct answers to each cue word. After they had provided the correct target word to a cue word twice, that word pair was dropped from the following test-with-feedback trials. On average, participants reached criterion for all word pairs after seven trials ($SD = 3$). Seven participants needed only three trials to reach criterion, and six participants required 17 to 24 trials to reach criterion for all word pairs. We observed no difference in the required numbers of trials to reach criterion between the 12 experimental groups.

In general, participants showed high memory performance in cued recall from the Initial Learning Session to Learning Sessions 2 and 3. Nevertheless, forgetting between sessions occurred as a function of the length of the ISI. At the end of Initial Learning, no differences emerged in cued recall between the three learning schedule conditions³ (contracting, $M = 95%$, $SD = 5%$; equal, $M = 94%$, $SD = 9%$; expanding, $M = 94%$, $SD = 8%$), $F(2,207) = 0.16$, $p = .849$, $\eta_p^2 = .002$.

³ Free-recall performance at the end of the Initial Learning Session also did not differ between the three learning schedules (contracting, $M = 53%$, $SD = 18%$; equal, $M = 51%$, $SD = 19%$; expanding, $M = 52%$, $SD = 18%$), $F(2,207) = 0.13$, $p = .875$, $\eta_p^2 = .001$.

In order to examine forgetting between learning sessions, cued-recall performances at the beginning of Learning Sessions 2 and 3 were used as dependent variable. Hence, we conducted a 2 x 3 repeated measures ANOVA with Session (Learning Session 2 vs. 3) as a within-subject

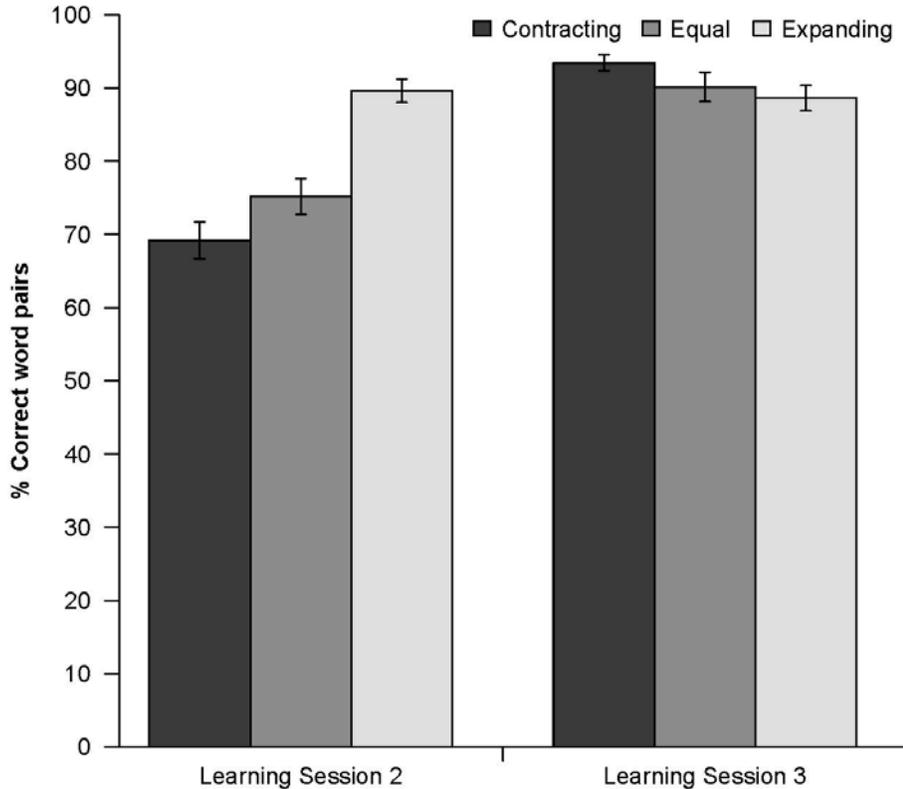


Figure 2. Percentages of word pairs recalled in the cued-recall tests at the beginning of Learning Sessions 2 and 3, as a function of learning schedule condition. Error bars represent *SEMs*.

factor and Learning Schedule (Contracting, Equal, or Expanding) as a between-subjects factor on memory performance at the beginning of the relearning sessions (Figure 2). We observed a significant effect of Session, $F(1,207) = 252.90, p < .001, \eta_p^2 = .55$: Participants retrieved more word pairs correctly on the cued-recall test at the beginning of Learning Session 3 ($M = 91\%$, $SD = 14\%$) than at the beginning of Learning Session 2 ($M = 78\%$, $SD = 20\%$). Moreover, a significant main effect of Learning Schedule was apparent, $F(2,207) = 5.23, p = .006, \eta_p^2 = .48$. Aggregated over Learning Sessions 2 and 3, participants maintained higher cued recall

performance in the expanding learning condition ($M = 89\%$, $SD = 14\%$) than in the contracting ($M = 81\%$, $SD = 14\%$), $t(136) = -3.30$, $p = .001$, $\eta^2 = .07$, or the equal ($M = 83\%$, $SD = 18\%$), $t(143) = -2.44$, $p = .016$, $\eta^2 = .04$, learning condition. Finally, we found a significant interaction between Session and Learning Schedule, $F(2,207) = 84.74$, $p < .001$, $\eta_p^2 = .45$. In line with the forgetting literature, cued recall at the beginning of Learning Session 2 showed a linear trend, by increasing from a contracting over equal to an expanding learning schedule⁴, $F(1,207) = 42.67$, $p < .001$. In contrast, and again in line with the forgetting literature, correct retrieval at the beginning of Learning Session 3 decreased in a linear fashion from a contracting over equal to an expanding learning schedule, $F(1,207) = 3.92$, $p = .049$.

Free- and cued-recall performances were assessed at the end of Learning Sessions 2 and 3 in order to measure memory after having restudied the material. Cued-recall performance was equally high in all learning schedule conditions at the end of both Learning Sessions 2 and 3, with performance $> 92\%$, $F_s \leq 0.93$, $p_s \geq .396$. Free recall performance at the end of Learning Session 2 also did not differ between learning schedules, $F(2,207) = 0.10$, $p = .906$, and was 54% on average. At the end of Learning Session 3, better free-recall performance was observed in the contracting condition than in the other two conditions, $F(2,207) = 4.17$, $p = .017$, $\eta_p^2 = .039$. A figure displaying both the free- and cued-recall performances at the end of Learning Sessions 2 and 3 is provided in Appendix A.

Final-test memory performance

Participants maintained high cued-recall performance for RIs of up to 7 days (i.e., final cued-recall performance of 91% and higher). Due to this ceiling effect in final cued recall, no

⁴ To revisit, the lags between the Initial Learning Session and Learning Session 2 were 5, 3, and 1 day(s) for the contracting, equal, and expanding learning schedules, respectively. The lags between Learning Session 2 and Learning Session 3 were 1, 3, and 5 day(s) for the contracting, equal, and expanding learning schedules, respectively.

significant effects of Learning Schedule or interaction with RI could be detected, with all F s ≤ 1.57 , p s $\geq .159$. Only a significant main effect of RI was found, $F(3,198) = 60.40$, $p < .001$, $\eta_p^2 =$

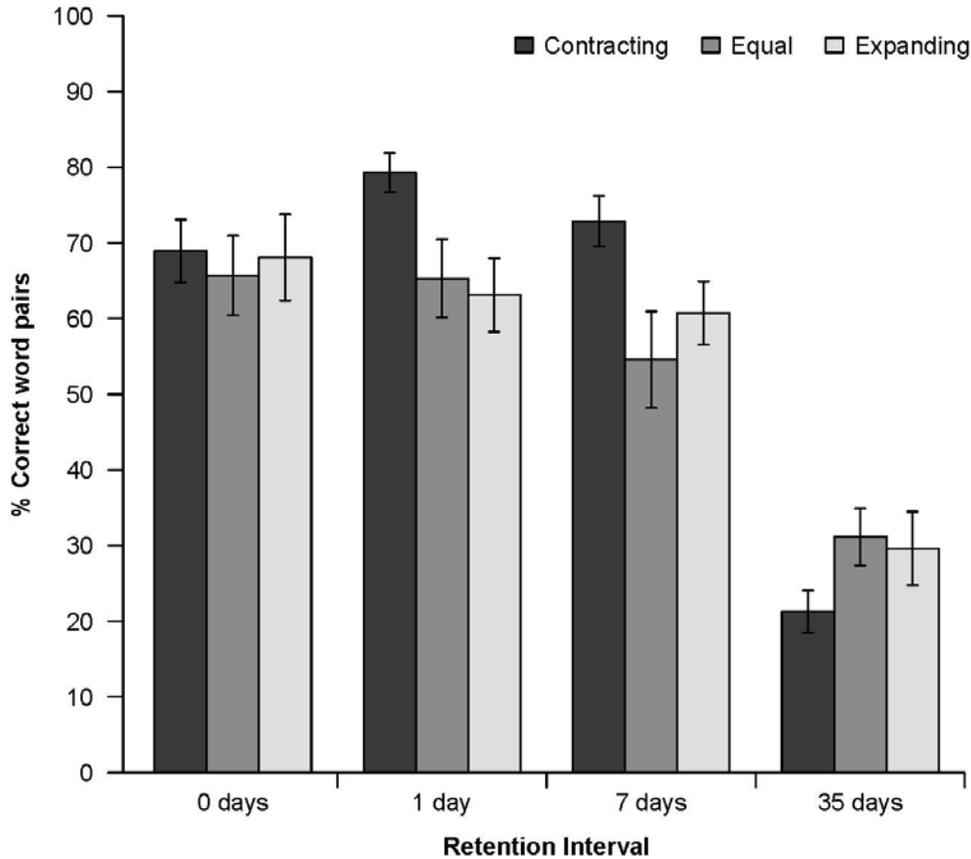


Figure 3. Mean percentages of correctly recalled word pairs on the free-recall test in the final test session, as a function of learning schedule and retention interval. Error bars represent *SEMs*.

.48. This effect was due to better cued recall performance after RIs of 0, 1, and 7 days ($M = 95\%$, $SD = 11$) than after an RI of 35 days ($M = 65\%$, $SD = 21$), $t(66.78)^5 = 10.24$, $p < .001$, $\eta^2 = .61$.

In the following analyses, we will focus on final free-recall performance, which was not at ceiling. Importantly, the overall result pattern of final cued-recall performance was similar to the pattern for final free-recall performance. For the sake of completeness, a graph displaying the final cued-recall performances is shown in Appendix B.

⁵ Whenever Levene's test indicated that the assumption of homogeneity of variances had been violated the degrees of freedom were adjusted accordingly to account for unequal variances.

For the free-recall test, items were considered correct if both words of a pair were recalled and correctly matched. A 3 (Learning Schedule) x 4 (RI) between-subjects ANOVA revealed a significant main effect of RI, $F(3,198) = 56.43, p < .001, \eta_p^2 = .46$: Participants recalled more word pairs after RIs of 0, 1, and 7 days ($M = 66\%, SD = 21$) than after an RI of 35 days ($M = 27\%, SD = 17$), $t(208) = 12.45, p < .001, \eta^2 = .43$. No main effect of Learning Schedule emerged, $F(2,198) = 2.06, p = .131, \eta_p^2 = .02$. However, a significant Learning Schedule x RI interaction did occur, $F(6,198) = 2.26, p = .040, \eta_p^2 = .06$. The results are visualized in Figure 3. We conducted planned comparisons separately for each RI condition and entered two contrasts that tested directly for the effects of interest. The first contrast tested the difference between the contracting condition and the other two learning schedule conditions, and the second contrast tested the expanding against the equal learning schedule condition⁶. In the 0-day RI condition, neither contrast led to a significant effect, $t(49) = -0.33, p = .747, \eta^2 = .002$, for the first contrast, and $t(49) = 0.33, p = .740, \eta^2 = .002$, for the second contrast. In the 1-day RI condition, the first contrast was significant, $t(47.18) = -3.43, p = .001, \eta^2 = .20$, indicating best free-recall performance for participants who practiced with a contracting learning schedule, as compared to the other two learning schedules. No significant difference was apparent between the expanding and equal learning schedules, $t(33.85) = -0.31, p = .761, \eta^2 = .003$. In the 7-day RI condition, we again found significantly better memory performance in the contracting learning schedule condition than in the other two learning schedule conditions, $t(41.22) = -3.00, p = .005, \eta^2 = .18$. Again, no difference emerged between the equal and expanding learning schedules, $t(29.39) = 0.80, p = .428, \eta^2 = .02$. Finally, in the 35-day RI condition, we found that participants in the contracting learning schedule condition performed worse than participants in the other two

⁶ We created two meaningful contrasts that were tailored to our hypotheses. The first contrast was contracting = -2, equal = 1, and expanding = 1, and the second contrast was contracting = 0, equal = -1, and expanding = 1.

learning schedule conditions, $t(48.13) = 2.21, p = .032, \eta^2 = .09$. No difference in free recall performance was detected between the equal and expanding learning schedule conditions, $t(34.60) = -0.27, p = .807, \eta^2 = .002$.

Discussion

Our data show that the optimal schedule for learning paired associates varies with the length of the RI. When the RI was 15 min (0-day RI condition), contracting, equal, and expanding learning schedules led to equivalent final free recall performance because the final tests followed immediately after the last learning session. This timing ensured good accessibility to the material that had recently been studied. Therefore, it is not surprising that – without a longer forgetting interval between the end of learning and final test – no effect of learning schedule emerged. In contrast, after the still-short RI of 1 day, the data showed a difference in final test performance as a function of learning schedule. We found that free-recall performance benefited more from a contracting learning schedule than from equal or expanding learning schedules for RIs of 1 day or 7 days. This contracting learning schedule superiority disappeared when participants were tested after a long RI of 35 days, and instead the equal and expanding learning schedules led to better final test performance. In no case did we find a difference between equal and expanding learning schedules, which always produced comparable memory outcomes.

We calculated Cohen's d effect sizes for the contracting learning schedule versus the combined expanding and equal learning schedules. These effect sizes ranged from large to very large (cf. Cohen, 1988), which emphasizes the potential importance of our results for real-world learning settings. More precisely, the effect sizes were $d = 1.55$ for the 1-day RI condition, $d = 1.16$ for the 7-day RI condition, and $d = 0.82$ for the 35-day RI condition. Put differently, in the

1-day RI condition, final free-recall performance increased by 23% when the material was studied in a contracting learning schedule rather than an equal or expanding learning schedule, and in the 7-day RI condition, the increase was 21%. Reversing this situation, in the 35-day RI condition, the increase for expanding or equal, relative to contracting, schedules was 43%.

Memory performance during learning was affected by the length of the ISI between learning sessions, with longer ISIs leading to more forgetting and lower cued-recall performance at the beginning of a relearning session. Correct cued recall at the beginning of a relearning session increased with the addition of learning opportunities. Thus, cued recall was higher at the beginning of Learning Session 3 than at the beginning of Learning Session 2. Participants benefited from relearning sessions and continuously improved their memory for the material. Our data confirm that overall, an expanding learning schedule indeed maintains higher performance during learning than does any other learning schedule (Landauer & Bjork, 1978). However, this benefit did not translate to generally superior final memory performance in our experiment (see also Logan & Balota, 2008).

Challenges to extant mathematical models

Our data challenge both ACT-R (Pavlik & Anderson, 2008) and MCM (Mozer et al., 2009) as models for the distributed practice effect with three learning sessions—at least with the parameter specifications used in Lindsey et al.'s (2009) simulation study. Lindsey et al. found that ACT-R predicts contracting learning schedules to be best—irrespective of the length of the RI. MCM predicts that the optimal learning schedule should vary with the RI, but in a different way. In MCM, the contracting learning schedule should not outperform the other two learning schedule types, regardless of RI. For long RIs (i.e., ≥ 1 day), the expanding learning schedule

should lead to better final memory performance than does the equal learning schedule. Our findings clearly contradict those predictions and call for a re-specification of MCM and ACT-R.

Our results can be accommodated best by the predictions made by the contextual variability theory as a stand-alone theory (Glenberg, 1979). For the shorter RIs of 1 day and 7 days, we showed a clear superiority for a contracting learning schedule. In accordance with the contextual variability theory, retrieval after RIs of 1 day and 7 days benefited from a greater overlap in contextual components between the components stored during the last two learning sessions and the components present during the final test. Consequently, for a long RI of 35 days, a higher matching probability between encoding and test context was obtained for an equal learning schedule, but also for an expanding learning schedule, and less so for a contracting learning schedule. The findings that equal and expanding learning schedules led to equivalent memory outcomes and the contracting learning schedule was inferior to them lead us to speculate that the ISI between Learning Sessions 2 and 3 plays a more important role in determining final memory performance than does the ISI between Initial Learning and Learning Session 2. It seems that contextual variability that is introduced by the later learning episodes affects final memory performance more so than contextual variability between the first two learning sessions.

Contextual variability theory provides a possible explanation for our unusual finding that for a 1-day RI, participants with a contracting schedule outperformed those in the 0-day condition on the final test, $t(26.32) = -2.12$, $p = .043$, $\eta^2 = .15$ (see Figure 3). In the 0-day condition, only the same-day learning session (i.e., Learning Session 3) would have provided strongly overlapping contextual cues with the final test. By contrast, for the contracting condition, the 1-day RI was more likely to benefit from retrieval cues from the previous two learning sessions.

It is possible that an adjustment of the model parameters of MCM could lead to more accurate model predictions, since it already incorporates assumptions of contextual variability

theory. A stronger emphasis on mechanisms suggested by the contextual variability theory, particularly in terms of weighing contextual components stored during the last two learning sessions more than components stored during the first learning session, might increase the ability of MCM to properly predict memory performance following three learning sessions.

Conclusion

In the present experiment, we investigated the optimal distribution of three learning sessions for the retention of paired associates after RIs of up to 35 days. This is the first demonstration that contracting intervals can outperform expanding and equal intervals when the RI is 1 day or 7 days. In contrast, when memory was assessed after an RI of 35 days, studying with equal or expanding intervals led to better performance than did contracting intervals. All effects were large in magnitude and led to educationally meaningful increases in test scores, making the effect of learning schedules a strong candidate that should be studied for generalization in less-controlled learning environments like schools. On the basis of the present findings, we advise planning review sessions while keeping the RI in mind. Teachers should make deliberate theoretically and empirically driven choices based on ideal implementation of the distributed practice effect, rather than relying on a simpler “utilize distributed practice” rule. In line with Cepeda, Vul, Rohrer, Wixted, and Pashler (2008), teachers’ choices become increasingly useful at maximizing learning as long-term retention becomes a more valuable educational goal. More specifically, we find that the best results for a test in 1 week can be obtained when the to-be-learned material is studied using a contracting schedule. However, if the goal is long-term accessibility of verbatim material (e.g., 1 month), one should plan learning sessions conforming to equal or expanding learning schedules.

Our findings show that the contextual variability theory can account best for the present results and that two extant mathematical models, ACT- R and MCM, will need to be revised to take the results of this study into account. Running multiday studies (such as the present experiment) and moving beyond a single methodology (e.g., the use of tests without feedback for expanding-interval studies) will stimulate further theory development that has potential to improve educational practice.

Besides strengthening the theoretical tie to empirical data, future research should examine the generalizability of our findings to more naturalistic learning environments and to more representative population groups (Henrich, Heine, & Norenzayan, 2010). Although our conclusions only apply to paired-associate learning, we have no reason to believe that they would not generalize to learning of other materials, although this issue should be addressed in future experiments. In terms of improving teaching practice today, it is clear that teachers' choices about when students relearn material greatly affects students' retention, both on an immediate test and in the long run, and high scores on an immediate test sometimes will be to the detriment of long-term retention.

Acknowledgments

The first author acknowledges support from the Ontario/Baden-Württemberg Faculty Mobility Program of the Ministry for Science, Research, and Arts of Baden-Württemberg, Germany, and from a postdoc fellowship from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). We are grateful to the Faculty of Health at York University in Toronto, Canada for supporting this project with a Minor Research Grant and an SSHRC Small Grant. We thank Masa Calic, Suzette Fernandez, and Ariella Winter for their assistance with data collection and Tina Weston for helpful comments on a previous draft of this article.

References

- Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, *120*, 20–33. doi: 10.1037/0096-3445.120.1.20
- Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering. Essays in honor of Henry L. Roediger III* (pp. 83–105). New York, NY: Psychology Press.
- 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
- 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., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619–636. doi: 10.1002/acp.1101
- 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, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- 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
- Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2*, 365–378. doi: 10.1037/1076-898X.2.4.365
- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Oxford, UK: Dover. (Original work published 1885)
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369–377. doi: 10.1037/h0046888
- 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, 65*, 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

- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. doi: 10.1017/S0140525X0999152X
- Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1250-1257. doi: 10.1037/a0023436
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*, 471–479. doi: 10.1080/09658210802647009
- Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, *38*, 116-124. doi: 10.3758/MC.38.1.116
- Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 704–719. doi: 10.1037/0278-7393.33.4.704
- Küpper-Tetzl, 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-Tetzl, C. E., Erdfelder, E., & Dickhäuser, O. (2013). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*. Advance online publication. doi: 10.1007/s11251-013-9285-2
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, R. N. Sykes, & the British Psychological Society (Eds.), *Practical aspects of memory* (pp. 625–632). London, UK: Academic Press.

- 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.
- Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition, 15*, 257–280. doi: 10.1080/13825580701322171
- 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.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3-8. doi: 10.1037/0278-7393.31.1.3
- Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*, 187–193. doi: 10.3758/BF03194050
- Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*, 101–117. doi: 10.1037/1076-898X.14.2.101
- Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27*, 431–452. doi: 10.1207/s15516709cog2703_5

- 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
- 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*, *10*, 30–39. doi: 10.1037/h0071614

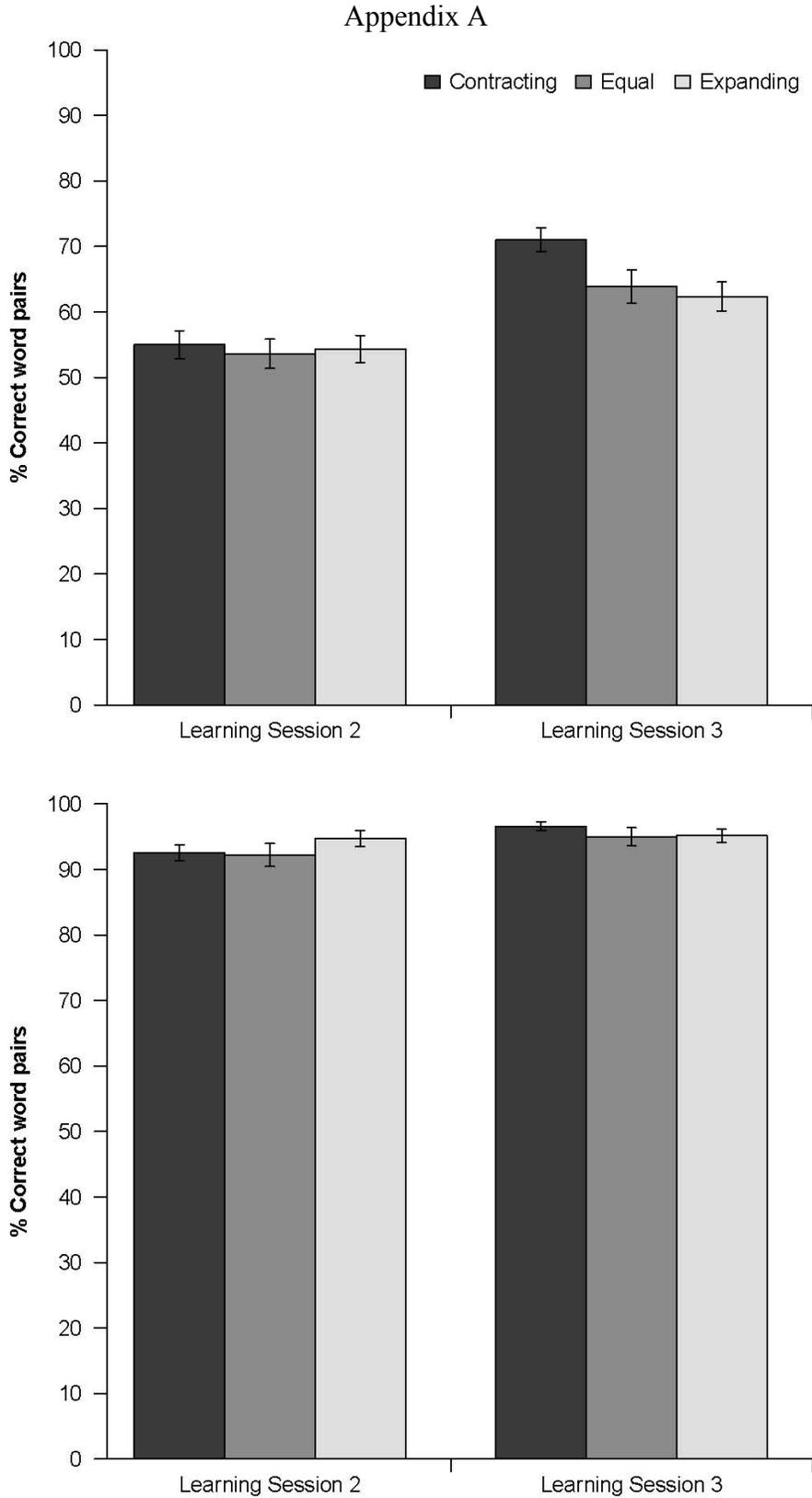


Figure 4. Mean percentages of correctly recalled word pairs on the free-recall (top) and cued-recall (bottom) tests at the ends of Learning Sessions 2 and 3, as a function of learning schedule. Error bars represent SEMs.

Appendix B

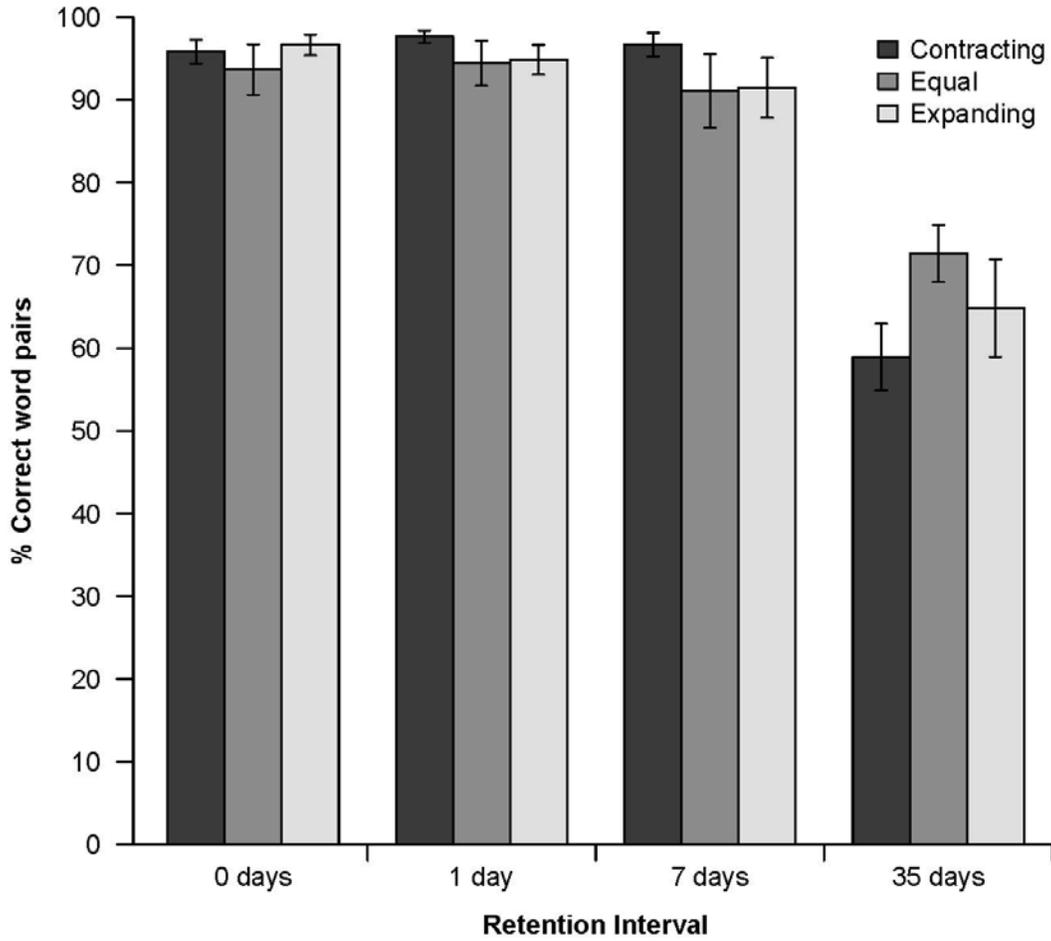


Figure 5. Mean percentages of correctly recalled word pairs on the cued-recall test in the final test session, as a function of learning schedule and retention interval. Error bars represent SEMs.