Title
The impact of practice frequency on learning and retention

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Tenison, Caitlin
Anderson, John R.

Publication Date
2017

Peer reviewed

The impact of practice frequency on learning and retention

Caitlin Tenison (ctenison@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University

John R. Anderson (ja+@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University

Abstract

The current study manipulated how frequently different problems were practiced during a first day of practice, with the more frequent items being more closely spaced. Fitting the data to a skill acquisition model, we find that greater spacing between items is associated with an increased probability of transitioning to more efficient phases of performance, but with a shallower speedup within each phase. Three days after training, we find that performance is predicted not by the practice frequency during training, but rather by the phase of skill acquisition attained during training. Thus, it is type of processing achieved not the amount and spacing of practice, that determines retention. Spacing, however, promotes learning by driving changes in cognitive processing.

Keywords: Skill acquisition; Practice frequency; Spacing effect; Learning; Retention

Introduction

A widely held observation, labeled the 'Spacing Effect', shows that increasing the time between practice opportunities improves retention. In contrast, massing practice opportunities together improves performance during training but negatively impacts retention. However, when the amount of time available for practice is limited spacing practice results in fewer total practice opportunities which may negatively impact learning (Anderson, 1982). Given the importance of the trade-off between spacing and amount of practice in understanding how humans learn, along with the clear educational benefits to improving retention, these effects have been studied extensively over the past 100 years. Explanation of both the effects of spacing and practice is critical for understanding skill acquisition. In this paper, we apply a model of skill acquisition (Tenison & Anderson, 2016) to a spaced-practiced data set to explore how spacing and the amount practice impacts the type of cognitive processes and affects the speed at which those processes are used to complete a task.

To capture the spacing effect, theories must explain both the short-term effects of practice frequency on skill acquisition and the long-term effects of spacing on retention. While several theories explain the locus of the spacing effect, we consider two qualitatively distinct sets of theories that study the processes occurring during practice that contribute to the spacing effect (Bjork & Allen, 1970). The set of deficit-processing theories focuses on the role of attention in the encoding of each practice opportunity. These theories hypothesize that, when practice opportunities are massed together, information used to encode and recall items is stored in working memory and little attention is needed to maintain these features. When practiced opportunities are spaced apart, features relevant to encoding are not stored in working memory and greater attentional resources must be applied to the maintenance and retrieval of memorized associations (Cuddy & Jacoby, 1982).

The second set of theories attributes the spacing effect to variation in encoding. These theories hypothesize that spacing between practice opportunities increases variability in contextual information used in the encoding of the item. When the item is seen again, the degree to which the item is encoded and associated with prior exposures is mediated by the time and number of items viewed between exposures (Landauer, 1969; Raaijmakers, 2003). This work explains why the benefit of spaced practice becomes apparent when retention is tested after a long period. An interesting phenomenon discussed and modeled by many encoding variability models investigates the increase in probability of recalling two different items, which have been seen only once, when those items were spaced further apart during training (Raaijmakers, 2003; Lohnas, Polyn, & Kahana, 2011). In this case, as with spacing, the probability of recalling both items at a much later time increases the more separated those items were when first seen.

The two types of theories view learning in terms of strengthening a memory rather than acquiring a skill. While this view may be sufficient for paired-associates task, it potentially over-simplifies the learning processes of more complex procedures. Given that the spacing effect has been observed across domains, including the learning of complex procedures (Shea, Lai, Black, & Park, 2000) and motor tasks (Shea et al., 2000), we stand to benefit from a closer consideration of how spacing and practice frequency influence procedural skill acquisition. Recent work by Tenison and Anderson (2016) suggests that skill acquisition is best described by both distinct changes in the cognitive processes used to perform a task, as well as quantitative improvements in the speed with which those cognitive processes are executed. With evidence from both behavioral and neuroimaging data, Tenison, Fincham, and Anderson (2016) suggest that these shifts follow the three phases of skill acquisition proposed by Fitts and Posner (1967). In the first phase, the Cognitive Phase, people must execute a series of procedures to perform a task. By the second phase, the Associative Phase, the response has been memorized and the task involves a single retrieval from memory. The third phase, the Autonomous Phase, the retrieval is dropped and the skill becomes a stimulus-response process. While this model has been operationalized in the ACT-R cognitive architecture (Anderson, 1982), prior work modeling the spacing effect within ACT-R predominantly accounts for the role of forgetting and memory rather than tracking changes in cognitive processing (Pavlik & Anderson, 2005). While this computational model has successfully modeled the impact of spacing on learning and retention, it is unclear how forgetting may interfere with skill acquisition.

In the current study, we apply an unsupervised learning approach developed to detect the phases of skill acquisition (Tenison & Anderson, 2016) to the problem solving latency generated in a spaced-practice task. In this task, participants are introduced to a novel mathematics operation and practice a set of problems. Different items receive different amounts of practice within the same learning period, resulting in greater spacing for the less frequently practiced items and less spacing between more frequently practiced items. Using a within-subjects design, participants were each exposed to 4 practice frequency conditions. We look at the effects of this different practice frequency manipulations on both learning during the task and retention three days later. We hypothesize that our manipulation will affect both the rate of achieving more advanced phases and the rate at which problem solving speeds up within a phase. We were interested in determining whether the phase of skill acquisition achieved during study and amount of practice within a phase would predict retention.

Methods

We ran our experiment on Amazon Mechanical Turk (MTurk). Participants signed up for two sessions, separated by a 66-72 hour period. On Day 1, participants were introduced to a novel math operator and practiced solving a fixed set of repeated problems. On Day 2 of the study, participants returned to complete the test including the items seen on Day 1 and answer questions about the strategies they used to solve the problems.

Participants

43 participants (15 female) completed both days of our experiment. All participants were from the United States. Participants represented a diverse age range, 20 to 53 years (M=30.9, SD=6.3), and education levels (highest level: 5 high school, 30 college, 8 graduate school). Reviewing the problem solving strategies participants had reported on Day 2, we excluded 8 participants who had used external aids on either day to solve problems. Our final sample included 35 participants. Participants were paid 2 cents for correctly solved problems (up to $8.60 on Day 1 and $3.60 on Day 2), as well as a $2 base pay for Day 1 and a $10 bonus for completing both days. This study was approved by the university internal review board and all participants gave informed consent for participation.

Materials and procedure

Participants learned a novel type of mathematics called a Pyramid problem. Pyramid problems follow the form of Base$Height, where the base indicates the first term in the additive sequence, and the height determines the number of terms to be added together (e.g., $8\$4 = 8+7+6+5$). Problem sets included heights of 3, 4, 5, and 6. The bases of our problem set varied from 4 to 11 with the restriction that the minimum base for a given height was height plus one. Because height determines the number of terms to sum, we use this as a means of manipulating problem difficulty. A total of 36 unique problems were used in the experiment. This study, however, will focus on only the 16 unique problems that were practiced on Day 1. For each problem, we recorded the accuracy and problem solving latency.

Day 1 After giving consent and completing a demographic information questionnaire, participants were introduced to the pyramid operation and given two blocks of 36 unique items to solve. Each item was presented on the screen in the form '8$4= _'. Following each input, we displayed corrective feedback by showing the expanded calculation and correct answer in the form '$8+7+6+5 = 26$'. After these pre-test blocks, participants then completed 10 practice blocks, which included 40 items each. During the practice period, participants practiced 16 unique problems. Practiced problems were divided into four Practice Frequency (PF) groups. Items in Practice Frequency-1 (PF-1) were seen once per block; in PF-2, twice per block; in PF-3, three times per block; and, in PF-4, four times per block. We included problems of four different heights (3-6) in each PF group, so that each block included 40 items total and 16 unique problems. By the end of Day 1, PF-1 items were seen 10 times, whereas PF-4 items were seen 40 times, as a result, our analyses are sensitive to both the effects of spacing and of general practice. Participants were given 4 hours to complete the tasks in Day 1.

Day 2 Participants were emailed a link to the retention test 66 hours after the initial completion of their Day 1 session. After the link was sent, participants had a total of 12 hours to begin the experiment (once started participants were limited to 2 hours to complete the experiment). Similar to the pre-test collected during the first two-blocks of Day 1, the retention test consisted of 10 blocks of 36 unique problems. This included 20 items that were only seen during the pretest on Day 1, and 16 problems from the four PF groups.

Results

The aim of this study is to explore the impact of spacing on the acquisition and retention of procedural skills. We divide our results into three sections. First, we report the general impact of spacing and practice frequency on learning and retention. We next fit the Tenison and Anderson (2016) to the data to generate parameter estimates and phase labels. Finally, we use mixed-effects modeling to explore the relationship between experimental condition and phase of skill acquisition.

Descriptive statistics

Before fitting a model to the data to identify learning phases, we examined the effect of spacing on the speed and accuracy of problem solving. A repeated measures analysis of

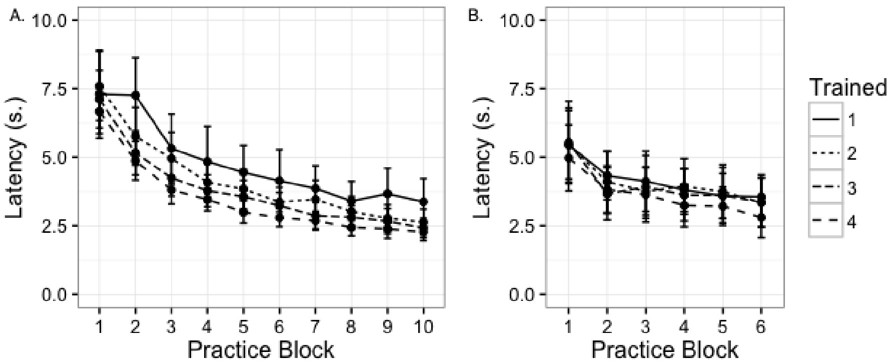

Figure 1: Problem solving latency for items averaged within each experimental block. Separate means are calculated for items of each practice frequency group. (a) indicates performance on Day 1 (b) indicates performance on Day 2. Error bars represent standard error

variance (ANOVA) run on mean, log transformed latency data revealed a significant main effect of practice frequency group ($F(9,306)=68.6$, $p<.001$) and block ($F(3,102)=34.2$, $p<.001$), and a significant interaction between PF and block ($F(27,918)=2.3$, $p<.01$). Figure 1 shows that response latency decreased across blocks and appears lower for higher-PF groups than lower-PF groups. The significant interaction suggests that the impact of block on speedup differs between PF groups. The average accuracy of items within the different PF groups also increases from PF-1 (M=92%, SD=.7%) to PF4 (M=97%, SD=.2%). A repeated measures ANOVA on accuracy data finds a significant main effect of PF group ($F(3,102)=8.7$, $p<.001$) such that PF-4 items were more accurate than lower practice frequencies, but the impact of block ($F(9,306)=.9$, $p=.5$) and the interaction between block and PF group ($F(27,918)=1.2$, $p=.3$) are not significant.

From the last items practiced on Day 1 to the first item practiced on Day 2, we see average decreases in accuracy from 95.6% (.5%) to 94 % (1.0%), and decreases in reaction time from 2.8 (.01 s) to 5.4 (.02 s). We will focus on latency for Day 2, which shows the large effect. A repeated measures ANOVA showed a significant main effect of practice frequency group ($F(3,102)=3.4$, $p<.05$) and problem difficulty ($F(3,102)=5.4$, $p<.005$) but no interaction ($F(9,306)=1.2$, $p=.3$). The effect of problem difficulty present in all groups suggested that on Day 2 many of these problems were solved rather than retrieved. Furthermore, the mean response times for these items indicated that items in the higher practice frequency groups were solved more quickly. These analyses show that while the effects present in Day 1 remain on Day 2, they are quite attenuated. However, as we will see this is because of the mixing of items that have reached different phases of learning on Day 1.

Model fitting

We fit the Tenison and Anderson (2016) power-law skill acquisition model to the response latencies for the items solved during the 10 practice blocks completed on Day 1. This model (refer to Tenison and Anderson (2016) for a detailed description) uses a Hidden Markov model (HMM) to track both the participants learning phase for any given problem and the number of practice opportunities a participant has had within a given phase. Using the within-phase tracking, we estimated parameters for a power-law function to describe speedup in the execution of the cognitive processes specific to each of the phases. However, according to the model, larger, abrupt changes are caused by transitioning to a more advanced phase of processing. We fit our model to each PF group separately. The model was fit separately for each item solved by a participant, but used trends across all participants solving items within a PF group in order to generate parameter estimates. We considered the number of phases that best fit the data by fitting HMMs with 1 through 5 possible phase transitions. Thus, we fit a total of 20 models (1 to 5 phases fit for each PF group). We used two measures to evaluate which model best fit the data: Bayesian Information Criterion, which penalizes models for added parameters, and log likelihood generated from a leave-one-subject-out cross validation. We best fit a 3-phase model for all 4 PF groups, replicating the result from our earlier studies. Once we determined the number of phases best fit by a model, we refit it to all the data and labeled each item with the phase the model identifies as most probable.

Because models are fit separately for each practice group, we first needed to establish that the 3 Phases identified by each model were in fact the same cognitive processes. In prior work, we found evidence that participants used calculation strategies in the first learning Phase and retrieval strategies in Phase 2 and 3 (Tenison & Anderson, 2016; Tenison et al., 2016). Because height determines the number of items participants sum together, we used the four different heights present in each PF group as a stand-in for problem difficulty. We would expect this effect of difficulty to disappear when participants use the retrieval strategies of Phase 2 and Phase 3. In Figure 2, each quadrant illustrates for each PF-group the effect of problem difficulty for the items in each Phase.

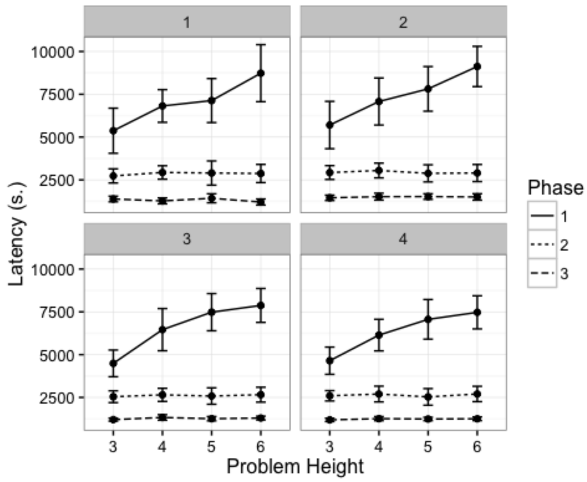

Figure 2: Mean problem solving latency for problems of different heights (i.e. difficulty levels) Each quadrant represents a different practice frequency group. Error bars represent standard error

We fit a mixed effects model to participants log latency data (Kliegl, Masson, & Richter, 2010) with random intercepts for each participant. Our model included an interaction between height and learning phase, along with a fixed-effect for training. We initially fit a maximal model in which we considered a random-effect for each fixed-effect in our model, and then used BIC to test whether or not removing those random effects improved the model fit (Barr, Levy, Scheepers, & Tily, 2013). Using this approach, we found our model was improved by inclusion of a random-effect for Phase, suggesting some variation across individuals in the impact of Phase on latency. We used the Kenward-Rodger (Kenward & Roger, 1997) approximation for degrees of freedom . In our final model, we could account for a significant amount of the variance in response time, with a fixed effect for the PF group of the item ($F(3,1194.8)=22.2$, $p<.001$) , problem difficulty ($F(3,1203.7)=43.9$, $p<.001$) and Phase ($F(2,44.9)=487.5$, $p<.001$), and a significant interaction between problem difficulty and Phase ($F(6,1200.3)=34.5$, $p<.001$). This analysis, as evident in Figure 2, shows that for all 4 practice frequency groups there is an effect of problem difficulty present in Phase 1, but not in Phase 2 or 3. The main effect of PF group, while difficult to discern from Figure 2, suggests a slight tendency for the more frequent items to be faster.

Interpreting model parameters

We gain insight into these effects by looking at the parameters of the 3-stage models that were estimated separately for each practice level. In this model, we fit a power function to each stage to reflect the within-stage practice. This is a 3-parameter function:

$$\mu_{ret} = I + \beta n^{-\alpha} \tag{1}$$

Where μ_{ret} is the time it takes to retrieve the answer, I is the asymptotic latency (i.e., the fastest possible time), β is the amount of latency that can be reduced with practice, n is the number of practice opportunities, and α is the learning rate. Asymptotic latency,I, and learning rate, α, parameters were estimated across all three Phases, and β was estimated separately for each Phase. These parameters capture the speed up within a Phase while our transition parameters, T_{12} and T_{23}, describe the probability of transitioning from one Phase to the next. Table 1 shows these parameters. Across models of the different practice-frequency groups the intercepts are essentially zero, implying that practice will always reduce the latency to some degree. The learning rate is controlled by the parameter a, which is small, indicating relatively small within-phase speedup. The parameters are remarkably similar across practice groups. There is a tendency for the speed up parameter to be greater for the higher-practice groups partially accounting for the faster within-phase times in Figure 1.

Table 1: Parameters for the three phase model of skill acquisition for each practice frequency group.

	I	α	β_{p1}	β_{p2}	β_{p3}	T_{12}	T_{23}
PF-1	3.7e-8	-.07	7.8	3.2	1.6	.17	.15
PF-2	9.5e-8	-.10	8.3	3.5	1.7	.13	.09
PF-3	3.5e-7	-.11	8.2	3.3	1.6	.07	.06
PF-4	5.9e-9	-.12	8.1	3.3	1.6	.06	.06

Because the number of practice opportunities vary between PF groups, the number of problems that reach Phase 3 is significantly different between the four groups ($F(3,102)=6.6$, $p<.001$), with 55.7% (6.7) of PF-4 problems reaching phase 3, 47.1% of PF-3 (6.1), 50% of PF-2 (5.7), and 35.6% of PF-1(6.7). However, the transition parameters, T_{12} (Phase 1 to 2) and T_{23} (Phase 2 to 3), are smaller for the high frequency group indicating that high frequency items spend more time in a phase before transitioning into the next Phase. On average, PF-4 items are seen 14.3 (1.1) times before transitioning to Phase 2 and 12.9 (1.0) before Phase 3, whereas PF-1 items are seen 4.7 (.29) times before Phase 2 and 3.6 (.25) times before Phase 3.

Retention after three days

To understand how the performance on Day 2 was impacted by practice on Day 1, we fit a mixed-effects model to the log latency of the first observation of each problem on Day 2. We explored fixed effects for practice frequency, Phase reached on Day 1, time spent within that Phase on Day 1, and problem difficulty. Fitting a maximal model with all factors and random effects, we systematically removed factors that did not account for variance within the model (see Barr et al. (2013) for method). In fitting our model, we found no improvement justifying the inclusion of random effects; and, we found that neither the practice frequency group nor the amount of practice within the last phase reached accounted

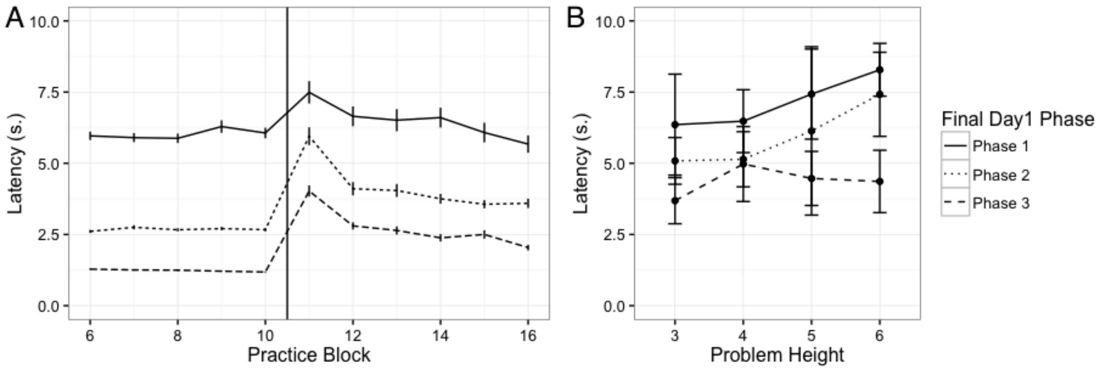

Figure 3: (a) To the left of the vertical line, mean problem solving latency for items in experimental blocks on Day 1; line types distinguish the mean latency for items in each Phase. To the right of the vertical line, mean problem solving latencies for items grouped by last Phase reached on Day 1 (b) Mean latency of first block of Day 2. Line type indicates Phase item reached on Day 1. Error bars represent standard error

for enough variance in response latency to justify either fixed effect. Our final model indicated a significant effect of problem difficulty ($F(3,515.8)=5.5$, $p<.001$), Phase reached on Day 1 ($F(2,542.5)=20.2$, $p<.001$), and a significant interaction between Difficulty and Phase ($F(6,519.1)=2.4$, $p<.05$). Figure 3a shows the mean effect of phase. It is striking that the items that were still in Phase 1 show little change in speed, while the items in Phases 2 and 3 slow down from Day 1 to Day 2. Figure 3b shows the mean response latencies for the first items solved on Day 2. The effect of problem difficulty appears to be present for both items that reached Phase 1 and 2 on Day 1, but less of an observable effect for items that reached Phase 3 on Day 1. Our interpretation of these results is that items in Phase 1 on Day 1 stay in Phase 1 and therefore show no changes in their latency patterns. However, some of the items in Phases 2 and 3 slip back a phase over the retention interval and therefore slow down. Items that slip back to Phase 1 will show a problem difficulty effect, possibly explaining the presence of a problem difficulty effect for items in Phase 2 (Figure 3b).

Discussion

Anderson and Milson (1989) suggest that memory phenomenon represent a joint function between general properties of memory and the strategies individuals use to process information. Our findings are aligned with prior studies that consider the impact of the spacing and practice frequency performance during training. Higher frequency in practice contributes to greater improvements in accuracy and response latency during training (e.g. Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The application of the skill acquisition model gives us insight into how problem solving strategies change in response to practice. We find the problem-solving latency advantage of massed trials is concentrated in the speedup within a phase rather than the transition between phases. This could be envisioned as 'rich get richer' process (Simon, 1955), in which learning strengthens both the probability of applying the previously used strategy and the speed with which the sub-procedures of that strategy are executed. Items that are spaced further apart exhibit shallower learning rates, which may make the search for a more efficient strategy more rewarding than the learning of problem solving sub-procedures. While theories of deficit processing or contextual variability could provide a mechanism for the differences in within phase speed up, the shift between phases of skill acquisition may be driven by a different mechanism.

Unlike prior work, which largely uses accuracy to measure the impact of spacing on retention, in our study we use response latency and the effect of problem difficulty. When we include information about what phase each item reached on the first day of training, we find that practice frequency no longer accounts for significant variance in problem solving latency at the retention test. Analyzing the speed of problem solving on Day 2, it appears that problems that items that reach Phase 2 and 3 on Day 1 are solved more quickly than items that remain in Phase 1. Additionally, the significant interaction between Phase and problem difficulty suggests that Phase 3 items may still be retrieved on Day 2, while Phase 1 and 2 items are calculated. This work is consistent with findings of Sisti, Glass, and Shors (2007) who found that the survival of neurons in the dentate gyrus and the strength of memory in an animal model was predicted not by whether or not practice was spaced or massed, but by how well the animals learned the task. While this study provides a biological mechanism for memory preservation, it is unclear computationally what memory process would explain the impact of phase on retention, but not on spacing nor general practice. In future work, we will explore how phase impacts retention by incorporating forgetting into our computational model of skill acquisition. In incorporating this capability, we will consider forgetting both in terms of regressing to a prior phase, and as regression within a phase. Additionally, including regression into the model will allow us to explore how spacing and skill acquisition on Day 1 impacts relearning on Day 2. In this future work we will limit the practice of more frequent items so to dissociate the effects of spacing from those of practice frequency. This work, while in an early stage, suggests that without considering the impact of skill acquisition on problem solving strategies, our models of the spacing effect, and memory more generally, are incomplete.

Acknowledgments

This work was supported by the National Science Foundation grant DRL-1420008, a James S. McDonnell Scholar Award, and by Carnegie Mellon University's Program in Interdisciplinary Education Research funded by the US Department of Education grant R305B090023. We would like to thank Shawn Betts for his help running the Mechanical Turk study.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 567–572.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, *132*(3), 354.

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, *21*(4), 451–467.

Fitts, P. M., & Posner, M. I. (1967). *Human performance.* Oxford: Brooks/Cole.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983–997.

Kliegl, R., Masson, M. E., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, *18*(5), 655–681.

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review*, *76*(1), 82.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of memory and language*, *64*(3), 249–255.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*(4), 559–586.

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the sam model. *Cognitive Science*, *27*(3), 431–452.

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*(5), 737–760.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425–440.

Sisti, H. M., Glass, A. L., & Shors, T. J. (2007). Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons. *Learning & memory*, *14*(5), 368–375.

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 749.

Tenison, C., Fincham, J. M., & Anderson, J. R. (2016). Phases of learning: How skill acquisition impacts cognitive processing. *Cognitive psychology*, *87*, 1–28.