

Avoidance of overlearning characterises the spacing effect

Doug Rohrer

University of South Florida, Tampa, FL, USA

The spacing of a fixed amount of study time across multiple sessions usually increases subsequent test performance—a finding known as the spacing effect. In the spacing experiment reported here, subjects completed multiple learning trials, and each included a study phase and a test. Once a subject achieved a perfect test, the remaining learning trials within that session comprised what is known as overlearning. The number of these overlearning trials was reduced when learning trials were spaced across multiple sessions rather than massed in a single session. In addition, the degree to which spacing reduced overlearning predicted the size of the spacing effect, which is consistent with the possibility that spacing increases subsequent recall by reducing the occurrence of overlearning. By this account, overlearning is an inefficient use of study time, and the efficacy of spacing depends at least partly on the degree to which it reduces the occurrence of overlearning.

Keywords: Spacing; Massing; Overlearning; Retention; Learning.

When a fixed amount of study time is distributed or *spaced* across multiple episodes, rather than concentrated or *massed* in a single episode, there is typically a boost in subsequent test performance known as the *spacing effect* (e.g., for a recent review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Indeed, the spacing effect is one of the most robust findings in the field of learning. What has largely gone unnoticed, though, is that spacing inherently reduces the occurrence of a second learning strategy that also reportedly boosts retention.

This second strategy is known as *overlearning*, and it occurs when a learner continues to study some particular set of material after reaching some criterion—typically one perfect trial—rather than defer any further

Correspondence should be addressed to Doug Rohrer, Department of Psychology, PCD 4118G, University of South Florida, Tampa, FL 33620, USA. E-mail: drohrer@cas.usf.edu

This research was supported by the Institute of Education Sciences of the US Department of Education (Grants R305H040108 and R305B070537). The opinions expressed are those of the author and do not necessarily represent the views of the Institute of Education Sciences or US Department of Education. Kelli Taylor provided invaluable assistance with data collection, and both Karl-Heinz Bäuml and an anonymous reviewer provided helpful comments.

study of this material to another session. For example, if a child repeatedly tries to recall the months of the year until all 12 months are recalled in the same attempt, any further *immediate* practice of this material constitutes overlearning. In previous experiments comparing an overlearning condition to one in which subjects quit after reaching criterion, overlearning typically improved subsequent test performance by a moderate margin (for a meta-analysis, see Driskell, Willis, & Copper, 1992). Thus, it is not surprising that overlearning is a widely advocated learning strategy. According to Hall (1989), for instance, "The overlearning effect would appear to have considerable practical value since continued practice on material already learned to a point of mastery can take place with a minimum of effort, and yet will prevent significant losses in retention" (p. 328). Similarly, Fitts (1965) concluded his paper by stating, "The importance of continuing practice beyond the point in time where some (often arbitrary) criterion is reached cannot be overemphasized" (p. 195).

However, although both spacing and overlearning increase subsequent test performance, the spacing of practice inherently reduces the amount of overlearning. This effect of spacing on overlearning was one of the results of the experiment reported here, but, as a means of providing an illustrative example for the remainder of this introduction, I will briefly describe a simpler, unpublished study from my laboratory that produced this same result. In this "pilot study", 20 learning trials were either massed in a single session or spaced across two sessions. (A third condition is not relevant here.) Each learning trial included a period of time devoted to the study of a list of paired associates, followed immediately by a test on the entire list. As these study-test cycles progressed, the proportion of items recalled on each successive learning trial, or *learning curve*, increased quickly at first and more gradually thereafter (Figure 1A). Two or six weeks after their last learning trial, subjects were given a final test without feedback, and a large spacing effect was observed at both test delays (Figure 1B).

The learning curves in Figure 1A illustrate how spacing affected overlearning. Both learning curves nearly reached the ceiling before the 10th trial, which means that most of the subjects within each condition achieved a perfect trial before the 10th trial. Therefore, most of the massers, who completed 20 *consecutive* trials, devoted all of their final 10 trials to overlearning. But spacers, who forgot many of the items during the 2-week spacing gap that separated their first 10 trials and their final 10 trials, had to devote at least some of their final 10 trials to reachieving the criterion of one perfect trial. Consequently, spacers completed fewer postcriterion trials than did massers, which means that spacing reduced the number of trials devoted to overlearning.

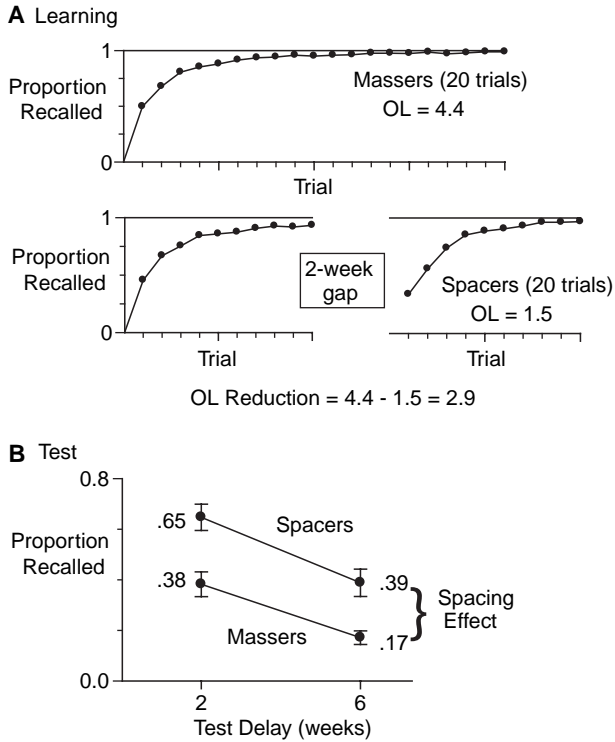


Figure 1. Results of pilot study. (A) Data represent mean proportion recalled during each learning trial. Once a subject achieved a perfect trial, the subsequent postcriterion trials within that session comprised overlearning, which was measured by the OL values shown earlier in the text (Eqn. 1). Because the spacing gap ensured fewer postcriterion trials for spacers than for massers, spacing caused an OL reduction (Eqn. 2). (B) Mean test scores revealed a large spacing effect at both test delays. Error bars represent one standard error.

MEASURING THE REDUCTION IN OVERLEARNING

Although learning curves visually illustrate how spacing reduces overlearning, a precise measure of this reduction requires a measure of overlearning. Herein, overlearning is measured in the usual fashion (e.g., Krueger, 1929):

$$OL = \frac{\text{Number of postcriterion learning trials}}{\text{Number of learning trials until criterion}} \quad (1)$$

For example, if a subject's first perfect trial occurred in the fifth of 20 consecutive trials, the subject completed 15 postcriterion trials, and $OL = 15/5 = 3$. With regard to the measure of overlearning for trials that are spaced across two sessions, I know of no precedent, and, thus, the OL value for two

spaced sessions is defined herein simply as the average of the OL values of each session (assuming that both sessions include the same number of trials, as is true throughout this paper). For example, if a subject achieved criterion on the fifth of 10 trials in the first session, so that $OL = 5/5 = 1$, and then reached criterion on the fourth of 10 trials in the second session, so that $OL = 6/4 = 1.5$, the overall OL value equals the average of 1 and 1.5, or 1.25.

Finally, and for the ease of exposition, the amount by which spacing reduces overlearning is defined here as the

$$OL \text{ reduction} = \text{Massers' OL} - \text{Spacers' OL} \quad (2)$$

In the pilot study, for instance, the Massers' OL (4.4) exceeded the Spacers' OL (1.5), and the OL reduction equalled the difference between these two values (2.9). In summary, the pilot study (and the experiment formally reported later) demonstrated two effects of spacing: it reduces overlearning (Figure 1A), and it increases test performance (Figure 1B).

OVERLEARNING REDUCTION HYPOTHESIS

These two effects of spacing—the reduction in overlearning and the increase in test performance—motivate the overlearning reduction hypothesis. By the weak form of this hypothesis, the amount of OL reduction (caused by spacing) *predicts* the size of the spacing effect. By the strong form, the amount of OL reduction (at least partly) *causes* the spacing effect. In brief, this hypothesis holds that the reduction in overlearning caused by spacing is not a *cost* of spacing but instead an advantage of spacing that might also directly contribute to its benefit.

In order to assess the relationship between the amount of OL reduction and the size of the spacing effect, the experiment presented here included a manipulation of both the schedule (spacing vs. massing) and the total number of learning trials (10 vs. 20). The overlearning reduction hypothesis predicts that the additional 10 learning trials should produce a concomitant increase in *both* the amount of OL reduction and the size of the spacing effect.

Although the effect of additional study time on OL reduction has not been previously investigated, its effect on the size of the spacing effect was assessed in an experiment by Metcalfe and Kornell (2003). In this study, each item was studied for 2 s or 16 s, and this duration was either massed as a single presentation or divided equally across two presentations spaced about half a minute apart. When tested a few minutes later, spacing and massing were shown to be equally effective for items studied for just 2 s each, but a large spacing effect was observed for items receiving 16 s of study. Thus, greater study time increased the size of the spacing effect.

The procedure for the experiment reported here differed from that used in the Metcalfe and Kornell (2003) experiment in two primary ways. First, the test delay was increased from a few minutes to 2 weeks, thereby providing a test of the generality of their finding. Second, and more importantly, subjects were repeatedly tested during the learning session (as in the pilot study), which provided a means of measuring the degree of overlearning.

METHOD

Subjects

The experiment was completed by 107 undergraduates at the University of South Florida. An additional 17 subjects completed the first session but failed to attend the second or third session, and these subjects were excluded from all analyses. None of the subjects participated in the pilot study described in the introduction.

Materials

Subjects learned to recall the name of the author when shown the title of a story, and every subject saw the following 10 title–author pairs: *Last Chance to See*—Adams, *Summer*—Camus, *The New Life*—Dante, *Exiles*—Joyce, *A Country Doctor*—Kafka, *A Winter Ship*—Plath, *The Seven Lady Godivas*—Seuss, *A Curious Dream*—Twain, *The Sphinx*—Wilde, *The Island of Statues*—Yeats. To verify that these title–author pairings were not known to the subject pool from which the subjects were drawn, an additional 50 subjects from the same subject pool were surveyed. In the survey, each subject was given the 10 story titles and asked to recall each author’s last name, and guessing was encouraged. None of the subjects provided any correct responses. Thus, although the author names might be familiar to native English speakers, the title–author pairings appear to be unknown to the subject population. Furthermore, any relevant preexperimental knowledge would not be a confound, because subjects were randomly assigned to conditions.

Procedure

All subjects attended three sessions, spaced 2 weeks apart. At the beginning of the first session, each subject was randomly assigned to one of four groups: high massers (20 massed trials), low massers (10 massed trials), high spacers (20 spaced trials), and low spacers (10 spaced trials). Spacers divided their learning trials across the first two sessions and were tested in the third session. Massers completed all of their learning trials in the first session and

were tested in the second session. In their third session, massers completed an unrelated task.

The first learning session began with a 60-s presentation of all 10 title–author pairs, and this was followed immediately by the learning trials. Each learning trial included a test and a study phase. During the 50-s test, each subject turned to a new page in a booklet listing the 10 story titles in a column and then, if possible, wrote each author’s name immediately to the right of the corresponding title. The order of the 10 titles varied randomly from trial to trial. In the immediately following 10-s study phase, subjects unfolded the right side of the booklet page, which caused each correct author name to appear immediately to the right of each subject’s response.

The final test was given 2 weeks after the last learning trial. Subjects were given a single page listing the 10 titles and asked to write the author name for each title. Three minutes were allotted, and no feedback was provided.

RESULTS

Learning

The proportion recalled on each learning trial, averaged across subjects, is plotted in Figure 2A. None of these learning curves reached the ceiling of 1.0, because each condition included at least one subject who never achieved one perfect learning trial. Yet, for the three conditions providing 10 consecutive learning trials (i.e., except for the low spacers), learning curves neared the ceiling by the 10th trial, indicating that most subjects achieved a perfect trial by the 10th trial. This meant that most of the high massers devoted all of their last 10 trials to overlearning. By contrast, the high spacers’ 2-week spacing gap led them to average just 36% on the first trial of the second session, which meant that they had to relearn the material. Consequently, the high spacers recalled significantly less than the high massers during their last 10 trials (83% vs. 96%), $t(52) = 4.26, p < .001$. The low massers underwent relatively little overlearning, though, as they quit after Trial 10. The low spacers, who divided only 10 trials across two sessions, underwent the least amount of overlearning, as their mean recall failed to surpass 90% on any trial. Consequently, the low spacers recalled significantly less than the low massers during the final five trials (60% vs. 89%), $t(51) = 5.71, p < .001$.

Overlearning reduction

As suggested by the shape of the learning curves, spacing reduced the OL values (Figure 2A), regardless of whether subjects completed 10 learning

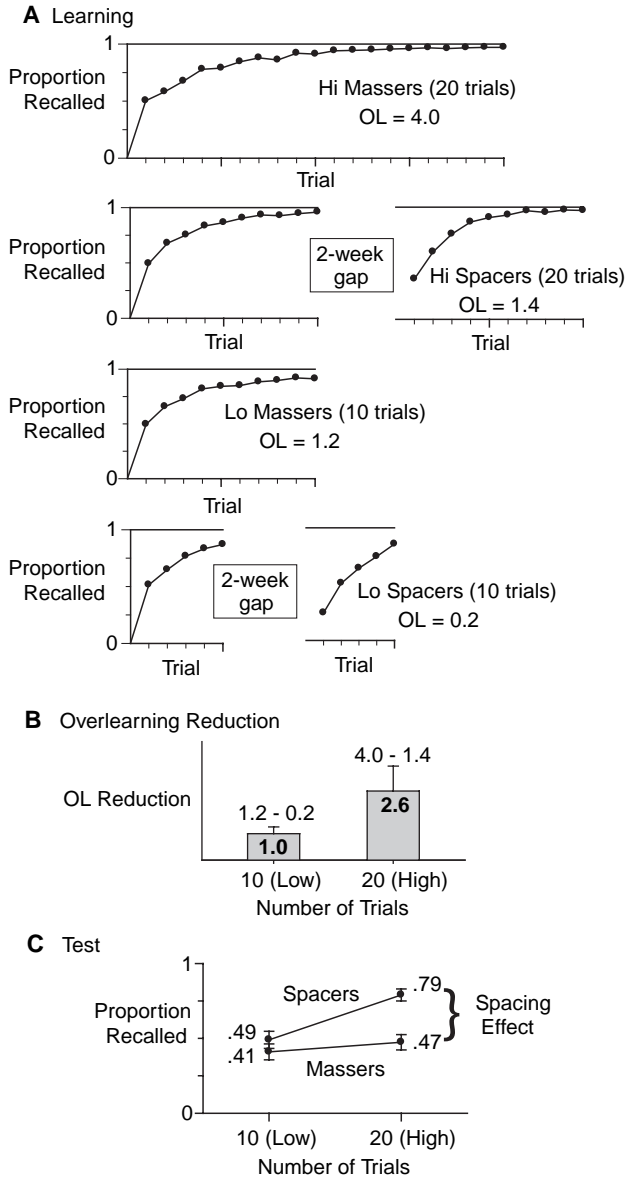


Figure 2. Results. (A) Data represent mean proportion recalled during each learning trial. Once a subject achieved a perfect trial, subsequent postcriterion trials within that session comprised overlearning, as measured by the OL values (Eqn. 1). (B) Spacing caused an OL reduction (Eqn. 2), and additional learning trials increased the amount of this OL reduction. (C) Spacing increased mean test scores, and the additional learning trials increased the size of the spacing effect. Error bars represent one standard error.

trials, $t(51) = 3.65$, $p < .001$, or 20 learning trials, $t(52) = 2.63$, $p = .01$. As a verification of this result, this analysis was redone by calculating an OL value for each *item*, as some researchers prefer. That is, rather than calculate a subject's OL value on the basis of the first trial in which *all* items were recalled correctly, an OL value for *each* item is calculated on the basis of when *it* was first recalled correctly. For example, if a subject correctly recalled the author Kafka on the second of 10 trials, the ensuing eight postcriterion trials would produce an OL value of $8/2 (= 4)$ for the item, Kafka, regardless of when other author names were first recalled. These item-based OL values are then averaged across items to obtain a single OL value for each subject. Item-level OL values are larger than list-level OL values, but, for the present data, at least, the pattern was unchanged: Spacing sharply reduced OL, regardless of whether subjects completed 10 learning trials (5.8 vs. 2.0), $t(51) = 11.68$, $p < .001$, or 20 learning trials (12.1 vs. 5.3), $t(52) = 9.08$, $p < .001$.

Effect of additional learning trials on OL reduction

Doubling the number of learning trials from 10 to 20 increased the amount of OL reduction from 1.0 to 2.6 (Figure 2B). Hence, the amount by which spacing reduced overlearning was greater if subjects completed 20 learning trials instead of just 10 learning trials. However, it is not clear how to assess the statistical significance of this increase because each value is itself a difference of two values (Massers' OL – Spacers' OL). Yet, because an increase in this difference is effectively a two-way interaction between schedule (massing vs. spacing) and learning trials (10 vs. 20), a two-factor analysis of variance was chosen. However, the two-way interaction was not statistically significant, $F(1, 103) = 2.40$, $p = .12$, apparently because of low power (0.2). However, when OL values were calculated at the item level, as explained earlier, the ensuing two-factor analysis of variance revealed a significant two-way interaction, $F(1, 103) = 13.63$, $p < .001$, $\eta_p^2 = .12$. Finally, although not relevant to the present discussion, both of the main effects in *both* of the analyses of variance were significant, all $ps < .001$.

Test

The increase from 10 to 20 learning trials sharply increased the size of the spacing effect (Figure 2C). A two-factor analysis of variance revealed a main effect of spacing, $F(1, 103) = 15.02$, $p < .001$, $\eta_p^2 = .15$, a main effect of more learning trials, $F(1, 103) = 12.84$, $p < .001$, $\eta_p^2 = .11$, and a two-way interaction, $F(1, 103) = 5.51$, $p < .01$, $\eta_p^2 = .05$. Subsequent Holm-Sidak pairwise comparisons revealed that high spacers significantly outscored high massers,

but low spacers and low massers did not differ reliably. Further Holm-Sidak tests showed that the increase from 10 to 20 learning trials significantly benefited spacers but not massers.

DISCUSSION

The primary finding reported here is that an increase in study time caused *both* an increase in the amount of OL reduction (Figure 2B) and an increase in the size of the spacing effect (Figure 2C). Specifically, whereas the spacing (rather than massing) of 10 learning trials produced only a slight reduction in the amount of overlearning and a statistically unreliable spacing effect, the spacing (rather than massing) of 20 learning trials produced both a large amount of OL reduction and a large spacing effect. Naturally, the usual caveats about generality apply, as it remains unknown whether the findings observed here would hold with different materials (e.g., a more abstract task), different procedures (e.g., recognition memory), and so forth.

As explained in the introduction, this finding is consistent with the overlearning reduction hypothesis, which holds that the degree to which spacing reduces overlearning predicts and perhaps contributes to the spacing effect. By this account, overlearning is an inefficient use of study time, and the efficacy of spacing depends at least partly on the degree to which it reduces the occurrence of overlearning. Hence, if the duration of a given period of massed study is too brief to allow overlearning, the spacing of this study time should yield a small spacing effect (or no spacing effect). By contrast, if a period of massed study is long enough to allow a large amount of overlearning, the spacing of this study effort should produce a large spacing effect.

The inefficiency of overlearning

More broadly, the inefficacy of postcriterion study trials suggests that, after a certain amount of massed study time, each immediately additional unit of study time devoted to the same material provides an ever smaller increase in the probability of later recalling that material. That is, although some minimal amount of study time might be needed to ensure at least some nonzero benefit, further *immediate* study provides diminishing returns. (More formally, for study durations exceeding some minimum amount of time, subsequent recall probability is an increasing, negatively accelerated function of study duration.) Hence, by this account, the additional study of some material is more useful when it is delayed rather than immediate, as was true in the present study.

Notably, this account holds that the benefit of additional, immediate study time diminishes *gradually*—not suddenly and sharply after the subject has achieved a particular measurable criterion (such as one perfect trial). Consequently, overlearning should provide *some* benefit, as was true in the present study (Figure 2C). Moreover, the benefit of, say, the first few postcriterion trials should be greater than the benefit of the next few postcriterion trials, and previous studies have shown this to be true (e.g., Krueger, 1929). In fact, this finding might partly explain why some overlearning studies find a greater overlearning effect than others, as some procedures fail to assess the effect of the initial amount of overlearning. In the present study, for instance, increasing the number of massed trials from 10 to 20 had only a small effect on subsequent test scores, but many subjects reached criterion *before* the 10th trial, as evidenced by Figure 2A.

By contrast, overlearning proved very beneficial in an experiment reported by Karpicke and Roediger (2008), and their procedure captured the effect of the initial overlearning. In this study, subjects learned 40 Swahili–English pairs with a procedure similar to the one used in the present study, and the number of test trials (e.g., *mashua*—?) within the learning session was manipulated. In one condition, subjects completed a fixed number of trials, as in the present study. But in the other condition, a subject who correctly recalled an item (e.g., *mashua*—“boat”) was never tested on that item during the remainder of the learning session. This “dropout procedure” eliminated all overlearning (i.e., $OL = 0$). One week later, subjects who used the dropout procedure recalled far fewer items than subjects who overlearned. (Incidentally, a dropout procedure could not be used in the present study, because it would have led spacers to complete more learning trials than massers, thereby confounding the study in favour of the spacing effect.)

However, although overlearning can boost test scores by a large margin, it must be remembered that this benefit is not without a cost, because, unlike spacing, overlearning requires *additional* study time. Thus, although overlearning might be more effective than not overlearning, as in the Karpicke and Roediger (2008) study, this does mean that overlearning is an efficient use of study time, as measured by subsequent recall probability *per unit of study time*. Indeed, the efficiency of postcriterion trials is typically worse than the efficiency of study trials preceding criterion (e.g., Krueger, 1929), as was observed in the present study. Hence, from a practical standpoint, the question is not whether overlearning is more effective than not overlearning, but whether overlearning is the most efficient use of the additional study time needed for overlearning. By the present findings, the answer to this question is clear: After reaching criterion, learners should avoid overlearning and instead shift their attention to other material. (There are, however, instances in which efficiency is less important than error-free

performance, such as the learning of certain medical procedures, and, in these cases, overlearning might be an ideal component of the optimal learning strategy.)

A boundary condition of the spacing effect

Finally, the present data also showed that a reduction in study time can reduce the size of the spacing effect, which, as described in the introduction, was shown previously by Metcalfe and Kornell (2003). In fact, in one of their studies, the spacing effect disappeared entirely when *total* study time per item was reduced to just 1 s. From a practical standpoint, this boundary condition suggests that the venerable strategy of spacing is subject to a critical caveat. Specifically, and as Metcalfe and Kornell also observed, a benefit of spacing might not be realised unless each of the spaced study periods is long enough to ensure that some level of mastery is achieved before attention is shifted elsewhere.

This caveat notwithstanding, the present data provide yet another demonstration of the benefit of spacing. Yet spacing remains underutilised by students (e.g., Bjork, 1979; Dempster, 1989; Rohrer & Taylor, 2006), even though the spacing effect has proven effective in classroom settings (e.g., Bahrick & Phelps, 1987; Bloom & Shuell, 1981; Metcalfe, Kornell, & Son, 2007; Seabrook, Brown, & Solity, 2005). One reason for this neglect of spacing might be the commonly held false belief that massing is superior to spacing (e.g., Kornell & Bjork, 2008; Son, 2004). Similarly, the widespread use of massing might stem partly from the belief that overlearning is an effective learning strategy, because massing rather than spacing increases the amount of overlearning. By the findings reported here, though, the reduction in overlearning caused by spacing is not a drawback of spacing but instead one of its benefits.

Original manuscript received June 2008

Revised manuscript received August 2008

First published online September 2008

REFERENCES

- Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over eight years. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 344–349.
- Bjork, R. A. (1979). Information-processing analysis of college teaching. *The Educational Psychologist*, *14*, 15–23.
- Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*, 245–248.

- 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.
- Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*, 309–330.
- Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*, 615–622.
- Fitts, P. M. (1965). Factors in complex skill training. In R. Glaser (Ed.), *Training research and education* (pp. 177–197). New York: Wiley.
- Hall, J. F. (1989). *Learning and memory*. Boston: Allyn & Bacon.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19*, 585–592.
- Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*, 71–78.
- Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*, 530–542.
- Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology, 19*, 743–768.
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209–1224.
- Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107–122.
- Son, L. K. (2004). Spacing one’s study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 601–604.