Educ Psychol Rev (2015) 27:635–643
DOI 10.1007/s10648-015-9332-4

REVIEW ARTICLE

Student Instruction Should Be Distributed Over Long Time Periods

Doug Rohrer [1]

Published online: 21 July 2015

Abstract In many academic courses, students encounter a particular fact or concept many times over a period of a few weeks and then do not see it again during the remainder of the course. Are these brief instructional periods sufficient, or should the same amount of instruction be distributed over longer periods of time? This question was the focus of several recent studies in which a fixed amount of instruction was distributed over time periods of varying duration and followed by a delayed posttest. With few exceptions, the results showed that longer instructional periods produced greater posttest scores if the posttest was delayed by at least a month or so. Notably, the search criteria for this review excluded several oft-cited studies favoring short foreign language courses over longer ones, but a closer look at these studies reveals limitations (e.g., no delayed posttest or confounding variables). In brief, the best reading of the data is that long-term learning is best achieved when the exposures to a concept are distributed over time periods that are longer rather than shorter.

Keywords Distributed · Spaced · Learning · Foreign · L2

Introduction

In a typical academic course, the exposures to a particular concept are distributed over a period of time that is much shorter than the duration of the course. For instance, many K-12 courses are divided into modules so that nearly all of the instruction on a particular topic, be it mitosis or Kepler's second law, is confined to a period of a few weeks even though the course lasts 10 months. In a similar scenario, the course itself is sometimes compressed into a shorter version, as when a traditional foreign language course is condensed into a few weeks so that students can receive the purported benefits of intensive learning. In both of these examples, a

✉ Doug Rohrer
drohrer@usf.edu

[1] Department of Psychology, University of South Florida, PCD4118G, Tampa, FL 33620, USA

fixed amount of instruction or practice is distributed over a time period that is shorter than the school calendar demands (Fig. 1). Should instruction be unnecessarily compressed in this fashion, or should the same number of exposures be distributed over longer periods of time?

The answer to this question might seem obvious because researchers have long extolled the benefits of distributed practice. Hundreds of studies have shown that students who *distribute* or *space* the study of a fact or concept over multiple sessions perform better on a delayed final test than do students who cram or *mass* the same effort into a single session— a finding known as the *spacing effect* (for recent reviews, see Bjork and Bjork 2011; Carpenter, Cepeda, Rohrer, Kang, and Pashler 2012; Cepeda, Pashler, Vul, Wixted, and Rohrer 2006; Dunlosky 2013; Dunlosky, Rawson, Marsh, Nathan, and Willingham 2013; Roediger and Pyc 2012; Schwartz, Son, Kornell, and Finn 2011; Son and Simon 2012; Willingham 2014; for various theoretical explanations, see Benjamin and Tullis 2010; Delaney, Verkoeijen, and Spirgel 2010; Toppino and Gerbier 2014.)

Yet, there are several reasons why educators might reasonably question the classroom relevance of the spacing literature. For instance, the superiority of spacing over massing has little practical significance in the classroom because students do not mass. Instead, any important fact or concept encountered in school is seen by students on at least two *different* days, meaning that all instruction is spaced. The relevant question for teachers is not whether spacing trumps massing but whether spacing over longer time intervals is superior to spacing over shorter time intervals. In other words, while it might behoove students to distribute a given amount of study over four class meetings rather than only one, this does not mean that the four class meetings should be distributed over a month rather than a week. The first question (four sessions vs. one session) differs *qualitatively* from the second question (four sessions over one week vs. four sessions over 1 month). To be sure, the latter question has been addressed in some studies, but properly controlled experiments of this kind are rare.

There are a number of other concerns about the ecological validity of the spacing literature. For instance, subjects typically complete the instructional period in less than an hour, and these brief learning phases tap memory processes that are unlike some of the processes employed by

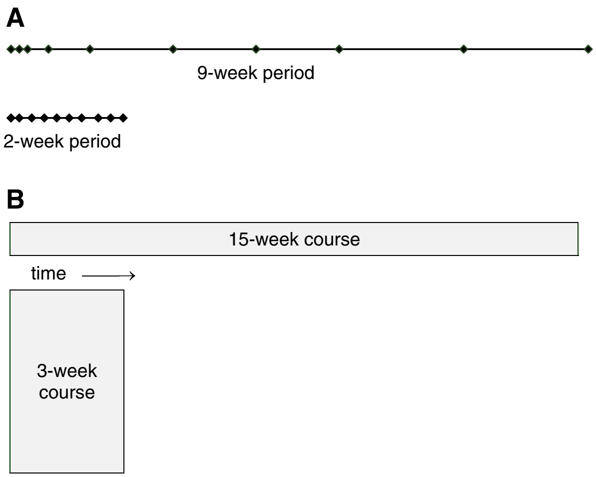

Fig. 1 Hypothetical comparisons of distributed practice. **a** Ten parabola problems can be distributed over 2 or 9 weeks. **b** A 45-h foreign language course can last 15 weeks or be condensed into 3 weeks

students who learn and relearn material over a period of days, weeks, or months. Another common criticism of the spacing literature is that most studies require only fact memorization (e.g., HOUSE-CASA), leading researchers to question whether spacing benefits higher level kinds of thinking such as inference and transfer (e.g., Donovan and Radosevich 1999; Kapler, Weston, and Wiseheart 2015).

Apart from any doubts about the practical relevance of the spacing literature, several studies of foreign language learning have assessed short and long versions of a foreign language course and found that the short course was *more* effective than the long course (for a review, see Serrano and Muñoz 2007). These results conflict with the findings reported in the spacing literature even though the studies comparing short and long foreign language courses are effectively studies of distributed practice (the same amount of instruction is distributed over a short or long time period). Can these two disparate sets of finding be resolved? The data supporting intensive language courses are not cited in the aforementioned reviews of the spacing literature, and these findings deserve a closer look.

For all of these reasons, researchers and practitioners can reasonably wonder whether the purported benefits of distributed practice extend to the classroom. Should teachers go to the trouble of distributing a given amount of instruction over longer periods of time? This question is the focus of the present review.

Evaluation Criteria

In order to better understand how student learning is affected by the duration of the instructional period, the present review is limited to studies satisfying the following criteria.

1. The study was published in an academic journal.
2. The study measured verbal or quantitative learning. This criterion excluded, for instance, studies of motor skills or music learning.
3. Student learning was measured by a final test, or *posttest*, and at least one final test was given after a delay of at least 1 day.
4. The study compared instructional periods of varying durations, and at least two of the instructional periods were longer than 1 day. For example, a study would be included if it compared instructional periods of 10 min, 10 days, and 10 weeks because two of the durations exceeded 2 days.
5. Students or classes were randomly assigned to condition.
6. The duration of the instructional period was not conflated with another variable. For example, in a study comparing short and long courses in foreign language, the two versions of the course must provide the same number of instructional hours. This criterion excluded a number of studies in which the duration of the instructional period varied systematically with test delay. For example, if one group studies on days 1 and 8, and another group studies on days 1 and 15, a final test on day 22 ensures that the latter group benefits from a shorter test delay.

These evaluation criteria excluded numerous studies of distributed practice. Many of the excluded studies were informative but were designed to answer questions other than the one examined here (e.g., Bahrick and Phelps 1987; Bloom and Shuell 1981; Gay 1973; Reynolds and Glaser 1964; Seabrook, Brown, and Solity 2005; Smith and Rothkopf 1984). However, the results of these excluded studies were the same as the results of the included studies, which is to say that the choice of criteria did not produce cherry-picked data.

Evaluation

A search of the literature turned up nine studies that met the evaluation criteria, and all were published since 2008. This set of studies is surprisingly small and astonishingly recent in light of the fact that the spacing literature consists of hundreds of studies dating back to the 1800s. The nine studies employed a variety of learning materials, procedures, and schedules (Table 1). The studies are summarized below, and a few are described in some detail.

The Studies

In an effort to assess whether spacing improves long-term retention in the classroom, Carpenter, Pashler, and Cepeda (2009) conducted a study in several eighth-grade US history classes. The experiment focused on material drawn from a course unit on slavery and sectionalism that was presented by the students' teachers during the final month of the school year (personal communication, S. Carpenter, October 20, 2014). The 1-month unit was followed by a researcher-created review session given 1 or 16 weeks later (by random assignment), meaning that the instructional period effectively lasted either 5 or 20 weeks. After the review, students waited another 9 months before taking a test consisting of factual questions (e.g., *Who assassinated President Abraham Lincoln? John Wilkes Booth*). Average test scores were about

Table 1 The studies reviewed here. In each experiment, total study time was fixed, but the study time was distributed over a period that varied. Distributed study over longer rather than shorter periods of time generally produced greater test scores when the test delay exceeded a month or so

Experiment	Material	Instructional period	Test delay
Bird (2010)	Foreign language	Five lessons distributed over 2 or 8 weeks (Fig. 2)	7 or 60 days
Carpenter et al. (2009)	History	One-month unit, followed by a review 1 or 16 weeks later	36 weeks
Cepeda et al. (2008)	Trivia	Two sessions separated by a gap of 0 to 105 days	1, 5, 10, or 50 weeks
Cepeda et al. (2009) Exp. 1	Foreign vocab	Two sessions separated by 0, 1, 2, 4, 7, or 14 days	10 days
Cepeda et al. (2009) Exp. 2	Trivia	Two sessions separated by 0, 1, 7, 29, 83, or 169 days	24 weeks
Kapler et al. (2015)	Science	One lecture, followed by a review 1 or 8 days later	5 weeks
Küpper-Tetzel and Erdfelder (2012)	Foreign vocab	Two sessions separated by 0, 1, or 11 days	1 or 5 weeks
Küpper-Tetzel et al. (2014)	Foreign vocab	Two sessions separated by 0, 1, or 10 days	1 or 5 weeks
Verkoeijen et al. (2008)	Nonfiction excerpt	Two readings separated by 0, 4, or 25 days	2 days

one third higher when the review was delayed by 16 weeks instead of 1 week, meaning that the longer instructional period proved superior.

Kapler et al. (2015) assessed whether this benefit of distributed practice holds for more abstract kinds of thinking. In this study, college students attended a simulated science lecture followed by a review 1 or 8 days later (by random assignment) and then returned after another 5 weeks for a test consisting of both factual questions and application questions. The students who waited longer before the review did better on both kinds of test questions.

A language learning study by Bird (2010) was designed to assess whether benefits of distributed practice depend on test delay. As part of an English language course for Malay-speaking adults, students practiced three kinds of verb tenses (simple past, present perfect, past perfect) by identifying and correcting previously-unseen, nongrammatical sentences (e.g., *I have seen the movie with my brother last week*, *Tony has quit his job last May*). The students practiced this task during five sessions spread evenly over a period of 2 or 8 weeks (by random assignment) followed 7 and 60 days later by a surprise final test consisting of novel sentences. Whereas the two groups of students performed about equally well on the 7-day posttest, average scores on the 60-day posttest were much higher for the group who distributed the sessions over the longer time period (Fig. 2).

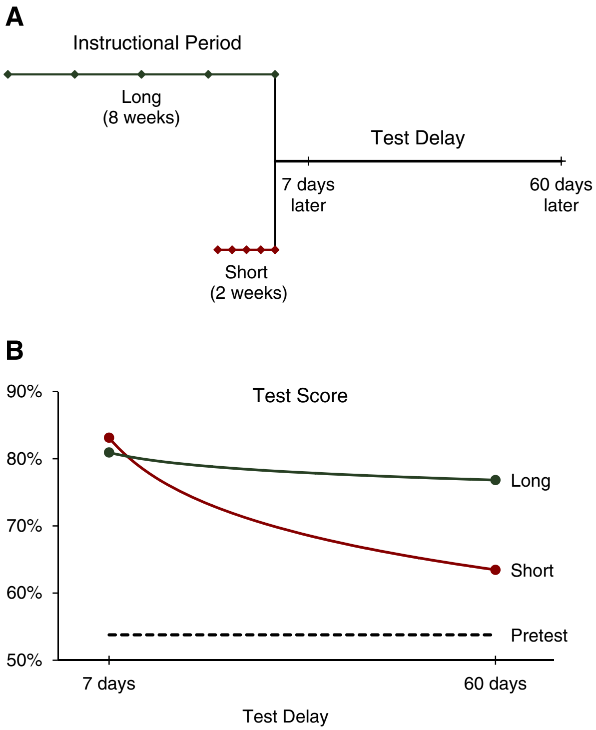

Fig. 2 Study by Bird (2010). **a** English-learning students practiced three kinds of verb tenses on five occasions distributed evenly over 2 or 8 weeks (by random assignment), and both groups of students were tested 7 and 60 days later. **b** On the test given 7 days later, the two groups of students scored about equally well (no statistically significant difference). On the test given 60 days later, the 8-week group outscored the 2-week group. Data points are means, and curves are hypothetical. The study included two kinds of tasks, and this graph shows data from one of the tasks (simple past vs. past perfect). The data for the other task showed the same pattern, but the 60-day test difference was slightly smaller

This futility of distributed practice after a brief test delay also was observed in a study reported by Verkoeijen, Rikers, and Özsoy (2008). College students in this study read a short excerpt from a nonfiction book and then reread the same text after a delay of 4 or 25 days. All students were tested only 2 days after the rereading, and the short delay group outscored the long delay group, though the difference was not statistically significant.

The relationship between test delay and distributed practice was the focus of the five remaining studies meeting the evaluation criteria, and all five studies used the same paradigm (experiments 1 and 2 in Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler 2009; Cepeda, Vul, Rohrer, Wixted, and Pashler 2008; Küpper-Tetzel and Erdfelder 2012; Küpper-Tetzel, Erdfelder, and Dickhäuser 2014). Students in each of these experiments studied a set of concrete facts during each of two study sessions (German-English pairs in the studies reported by Küpper-Tetzel and her colleagues, and trivia facts in the other three studies). The two sessions were separated by gaps as long as 169 days and followed by test delays up to 50 weeks (Table 1). Although the findings of these studies are rather complex, the takeaway is that the spacing gap that produced the highest test scores was *not* always the longest gap. Instead, the optimal spacing gap depended on the test delay, with longer test delays demanding increasingly longer gaps. For example, in the study by Cepeda et al. (2008), students were tested after a delay of 1, 5, 10, or 50 weeks, and students in *each* group were randomly assigned to a particular spacing gap (the study included 1354 subjects) Among students who were tested after a delay of 5 weeks, the best scores were achieved by students whose two sessions were separated by a gap of 11 days, with both shorter and longer gaps proving less effective. Among the students who were tested after a delay of 10 weeks, students did best when their two study sessions were separated by 21 days.

However, although these studies found that longer test delays required increasingly longer spacing gaps, this finding also means that a spacing gap was sometimes *too* long—even when the test delay was long. For instance, in the study by Cepeda et al. (2008), students who were tested after a delay of 50 weeks scored higher on the test if their two study sessions were separated by a gap of 70 days rather than 105 days. It is not clear why a gap might be too long, though one possibility is that the spacing effect dissipates when the gap between two sessions is long enough for students to have forgotten what they learned in the first session. However, even by these data, the test score penalty for a too-short gap was much greater than the penalty for a too-long gap.

More importantly, the finding that a spacing gap can be too long does not limit the duration of an instructional period because a long instructional period can be composed of brief spacing gaps. For example, if students see a particular concept on five different days during an instructional period lasting 4 months, each of the spacing gaps between any two *successive* exposures can be shorter than 4 weeks. In brief, these data are consistent with the claim that longer instructional periods are better than shorter ones.

Summary

The studies reviewed here required students to study a variety of learning materials over instructional periods lasting days, weeks, or months, and the results of these studies generally tell a rather simple story. Distributing a given amount of study time over a longer rather than shorter period of time leads to increased posttest scores if, and perhaps only if, the test is given after a delay of at least a month or so. Put another way, longer instructional periods generally produce more durable learning.

The Foreign Language Studies

Whereas the data reviewed here favor longer periods of instruction, several studies of intensive foreign language instruction have found that shorter courses are *more* effective than longer ones. In light of this discrepancy, these language learning data are briefly described here even though these studies did not meet the evaluation criteria. It turns out that this literature has limitations that cloud its interpretation.

Many of the studies favoring intensive language learning compared short and long courses that differed in ways other than their duration. For instance, one group of researchers has reported several studies showing that an intensive English program for French-speaking Canadian children was more effective than the traditional longer program (e.g., Collins, Halter, Lightbown, and Spada 1999), but it appears that the shorter program provided more instructional hours than did the longer program. Furthermore, Collins et al. note that the short program was "limited to students who had above-average academic ability" whereas the long program included students with a "wide range of ability" (p. 660). This confounding of ability and course duration cannot be addressed by measuring the improvement between a pretest and posttest because students of greater ability tend to be more motivated as well, and greater motivation translates to greater gains between pretest and posttest.

Among the nonconfounded studies favoring intensive language learning, it appears that none included a *delayed* posttest. For instance, Serrano and Muñoz (2007) and Serrano (2011) assessed the learning of adult Spaniards in a 110-h English course spanning 1 or 7 months and found that the shorter course produced better learning on a variety of measures (with some exceptions), but all of the learning measures were collected during the course. It therefore remains unknown whether the students in the longer course would have produced higher scores on a test given at least a month or so later, which is exactly what was found in the Bird (2010) study of foreign language learning (Fig. 2). Indeed, as Serrano and Muñoz (2007) conclude in their final paragraph, "Further research should include delayed posttests in order to observe how much knowledge is retained after some time has passed…" (p. 319). In brief, a short intensive language course might be ideal for people preparing for an upcoming trip to Paris, but the data do not suggest that the shortening of a language course provides more durable learning.

Practical Implication

The data support the use of longer instructional periods, and this finding can be exploited by instructors in a number of ways. In some cases, this can be accomplished by altering students' assignments and activities. For instance, one popular vocabulary workbook series used in grades K-12 is divided into dozens of lessons, and each lesson consists of several activities devoted to a set of words not seen in the other lessons. Rather than have students complete all of the activities in each lesson before moving on to the next session, the activities can be reordered so that the students encounter the same set of words several times over several months rather than several times in a single week.

To be sure, some kinds of material are presented more coherently when the material is divided into discrete units (e.g., a biology course is often divided into units on genetics, plants, human physiology, and so forth), but this modular organization does not necessarily demand that the exposures to any particular fact or concept be confined to its unit. In some courses, for instance, students might be asked to synthesize previously learned material (e.g., "List similarities and differences between humans and plants," or "Compare and contrast the American Revolutionary War and the American Civil War"). Instructors also can give cumulative exams with answer feedback. For instance, teachers might add to each exam a few questions on material learned much earlier in the course and then make sure that students correct their mistakes. To be clear, though, none of these recommendations have been empirically evaluated, and more research is needed.

In mathematics and quantitative science courses, teachers can lengthen instructional periods by assigning sets of practice problems that provide what is known as interleaved practice. Traditionally, sets of practice problems consist solely of problems relating to the same concept or procedure (e.g., a group of problems requiring students to add fractions, or a group of problems requiring students to solve an equation using the quadratic formula), and the drawback of this approach is that students know the relevant strategy for each problem *before* they read the problem. This scaffolding provides an illusion of mastery because students must know how to choose an appropriate strategy when they encounter a problem on a cumulative exam or standardized test. With interleaved practice, problems of different kinds are mixed together within the same assignment so that students must learn to choose an appropriate strategy on the basis of the problem itself. Interleaved assignments are created by rearranging the practice problems in the course so that different kinds of problems appear in the same assignment, and this intrinsically guarantees that problems of the same kind are distributed across many assignments, thereby increasing the duration of the instructional period. A handful of studies have found that interleaved practice boosts test scores, especially after longer test delays (e.g., Rohrer, Dedrick, and Stershic 2015).

Finally, it is worth repeating that the benefits of distributing instruction over longer periods of time hold only when students are tested after a delay of a month or so, though the critical delay almost certainly depends on both the type of material and the total amount of instructional time devoted to the concept. In fact, the data reviewed here suggest that scores on immediate tests are *reduced* when instruction is distributed over longer periods of time, and this boundary condition has been observed in numerous studies that did not happen to meet the inclusion criteria used in this review (e.g., for example, Rawson and Kintsch 2005; for a review, see Cepeda et al. 2006). Although this caveat presumably matters little to teachers who want their students to achieve durable learning, it means that students who cram all of their study effort into a period shortly before an exam are behaving more rationally than it might first seem (e.g., Bjork, Dunlosky, and Kornell 2013). This cramming, though, is not without its costs. As the data show unequivocally, material covered more quickly is more quickly forgotten.

Acknowledgments This work was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110517. The opinions expressed are those of the author and do not necessarily represent the views of the U.S. Department of Education. I thank Kelli Taylor Zarate, Andrew Butler, Shana Carpenter, and three anonymous reviewers for their suggestions.

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.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228–247.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31*, 635–650.

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

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444.

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.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U. S. history facts. *Applied Cognitive Psychology, 23*, 760–771.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing toenhance diverse forms of learning: review of recent research and implications for instruction. *Educational Psychology Review, 24*, 369–378.

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.

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, 11*, 1095–1102.

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

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly, 33*(4), 655–680.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation, 53*, 63–147.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*, 795–805.

Dunlosky, J. (2013). *Strengthening the student toolbox: Study strategies to boost learning* (pp. 12–21). Fall: American Educator.

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

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

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction, 36*, 38–45.

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: a multinomial processing tree analysis. *Memory, 20*(1), 37–47.

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*(3), 373–388.

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology, 97*, 70–80.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55*, 297–308.

Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242–248.

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*

Schwartz, B. L., Son, L. K., Kornell, N., & Finn, B. (2011). Four principles of memory improvement: a guide to improving learning efficiency. *International Journal of Creativity and Problem Solving, 21*, 7–15.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: from laboratory to classroom. *Applied Cognitive Psychology, 19*, 107–122.

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning, 61*(1), 117–145.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: any difference in EFL? *System, 35*, 305–321.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*, 341–358.

Son, L. K., & Simon, D. A. (2012). Distributed learning: data, metacognition, and educational implications. *Educational Psychology Review, 24*(3), 379–399.

Toppino, T. C., & Gerbier, E. (2014). About practice: repetition, spacing, and abstraction. *The Psychology of Learning and Motivation, 60*, 113–189.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Özsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology, 22*, 685–695.

Willingham, D. T. (2014). Strategies that make learning last. *Educational Leadership, 72*(2), 10–15.