Learning and Instruction 36 (2015) 38–45

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning

Irina V. Kapler*, Tina Weston, Melody Wiseheart

York University, Department of Psychology, LaMarsh Centre for Child and Youth Research, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

A R T I C L E I N F O

Article history:
Received 11 November 2013
Received in revised form
7 November 2014
Accepted 8 November 2014
Available online

Keywords:
Spacing effect
Classroom
Long-term memory
Retention
Higher-level learning

A B S T R A C T

Despite showing robust benefits in lab-based research, there remain relatively few studies exploring the spacing effect in educational contexts with meaningful materials. In this study, participants ($N = 169$ undergraduate students) attended a simulated university lecture where they were presented with natural science curriculum material. Participants reviewed the material either one day or eight days after the lecture via an online review. Participants completed a final test on the material five weeks after each respective review. During the review and final test participants were asked both factual and higher-level (application) questions. Results showed that reviewing material eight days after the lecture led to better final test performance for both types of questions when compared to reviewing only one day after the lecture. This study suggests that spaced review is a robust and effective strategy that can be and should be adapted to classroom practice.

1. Introduction

The spacing effect – a memory advantage that occurs when study sessions are spaced apart in time – is a widely recognized phenomenon in cognitive psychology. It is known to improve long-term retention and thus has clear implications for educational settings (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Pashler, Rohrer, Cepeda, & Carpenter, 2007). However, few experiments have explored the benefits of spacing in actual classroom settings with real-world curriculum and educationally relevant inter-study and retention intervals (Dempster, 1988). Therefore, it might be premature to suggest that spacing be systematically implemented in real-world classroom environments. This study introduces a spacing manipulation into a simulated classroom, using undergraduate curriculum material, educationally relevant inter-study intervals of 1 day and 8 days, and a retention interval of 35 days.

1.1. The spacing effect in the laboratory

To date, hundreds of studies have demonstrated that spaced learning reliably and robustly improves long-term retention across a wide variety of laboratory-based memory tasks (for reviews see Cepeda et al., 2006; Donovan & Radosevich, 1999; Janiszewski, Noel, & Sawyer, 2003). For example, it has been found to benefit name learning (e.g., Carpenter & DeLosh, 2005), object learning (e.g., Cepeda et al., 2009), vocabulary learning (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Kornell, 2009), fact learning (e.g., DeRemer & D'Agostino, 1974), text passages (e.g., Gordon, 1925; Rawson & Kintsch, 2005; Verkoeijen, Rikers, & Ozsoy, 2008), mathematical concepts (e.g., Rohrer & Taylor, 2006; 2007), motor skills (e.g., Baddeley & Longman, 1978; Mackay, Morgan, Datta, Chang, & Darzi, 2002; Moulton et al., 2006; Panchuk, Spittle, Spittle, & Johnston, 2013; Shea, Lai, Black, & Park, 2000), and musical skills learning (e.g., Simmons, 2012).

1.2. The spacing effect and higher-level learning

The majority of spacing studies, however, have used verbal or factual material as the to-be-learned stimuli (e.g., 839 effect sizes reported in Cepeda et al., 2006), where participants are not required to do anything with the information other than retrieve it from memory (Moss, 1995). Yet, in educational settings, it is rare that students are required to simply retrieve isolated pieces of information from memory; instead, they often must manipulate and apply the remembered information to answer more complex, higher-level questions. In comparison to hundreds of factual material spacing studies, the number of studies examining the effect of

* Corresponding author. Tel.: +1 416 736 2100x44037; fax: +1 416 736 5814.
E-mail addresses: tchertok@yorku.ca (I.V. Kapler), westont@yorku.ca (T. Weston), ncepeda@yorku.ca (M. Wiseheart).

http://dx.doi.org/10.1016/j.learninstruc.2014.11.001
0959-4752/

spacing on higher-level learning is limited. Some of the higher-level skills examined so far include learning of mathematical (e.g., Gay, 1973; Pyle, 1915; Rohrer, 2009; Rohrer & Taylor, 2006; 2007) and science concepts (e.g., Gluckman, Haley, & Sandhofer, 2014; Reynolds & Glaser, 1964; Vlach & Sandhofer, 2012), inductive category learning (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Vlach, Sandhofer, & Kornell, 2008; Wahlheim, Dunlosky, & Jacoby, 2011; Zulkiply & Burt, 2013; Zulkiply, McLean, Burt, & Bath, 2012), and the ability to make complex judgments (e.g., Helsdingen, van Gog, & van Merriënboer, 2011). These studies demonstrate that the spacing effect exists not only for simple fact learning but also for the learning of more complex material. Despite this more recent accruement of evidence, studies examining the effect of spacing on higher-level learning are few. Ecologically valid studies that use educationally relevant materials, timescales, and methods are needed before specific recommendations to educators can be made (e.g., Dempster, 1988; Dunlosky & Rawson, 2012).

1.3. The spacing effect in the classroom

This study contributes to the growing literature exploring spacing effect benefits in applied settings. Currently, there are some studies examining the spacing effect in the classroom (e.g., Balch, 2006; Bird, 2010; Bloom & Shuell, 1981; Carpenter, Pashler, & Cepeda, 2009; Fishman, Keller, & Atkinson, 1968; Küpper-Tetzel, Erdfelder, & Dickhäuser, 2014; Seabrook, Brown, & Solity, 2005; Smith & Rothkopf, 1984; Sobel, Cepeda, & Kapler, 2011; Yazdani & Zebrowski, 2006; see Kiepert, 2009, for a review). One possible reason for the lack of spacing effect classroom studies, in comparison to hundreds of laboratory studies, is the abundance of extraneous variables (noise) present in the classroom that can affect the success of a spacing intervention (e.g., classroom peer distractions, students' previous knowledge of the subject, high attrition rates, class schedule-induced time constraints, etc.). In comparison, many laboratory studies use computerized paradigms (e.g., a word pair presented on a screen for a specified number of seconds), where participants are tested individually, their attention is directed to the computer screen, and they are often required to learn material to some criterion before advancing to the next part of the experiment. In the classroom, however, students are part of a larger group of peers, the mode of delivery of the lesson is the decision of the teacher (often a lecture format), and students' attention may be diverted for any number of reasons, placing into question how well the information is initially understood. When one considers these complexities of classroom practice, it is easy to see how the benefits of spacing may be weaker in the classroom than in a controlled laboratory setting.

Many existing classroom studies use simple verbal or factual material as stimuli. For example, Carpenter et al. (2009) looked at the effects of testing and spacing in 8th-grade students learning U.S. history facts. After being taught in class, history facts were reviewed after 1 week, 16 weeks, or not at all, with the final test taking place 9 months later. Review took the form of testing followed by feedback or simply re-reading the facts. The study found significant testing effect benefits and spacing effect benefits approached significance with $d = 0.5$, $p = .06$. Sobel et al. (2011) taught fifth-graders the definitions of eight GRE vocabulary words. Half of the words were reviewed immediately, and half one week later. After a five week retention interval, words reviewed one week after initial learning showed spacing effect benefits with $d = 0.5$, $p = .004$. Bloom and Shuell (1981) divided a high school French class into two learning groups. One group learned 20 French words in three 10-min sessions spaced over three consecutive days. The other group learned the same 20 French words in a single 30-min session massed in one day. In a final free-recall test four days after learning the words, students in the spaced learning group outperformed students in the massed learning group, with $d = 1.0$, $p < .01$ (see also Balch, 2006; Fishman et al., 1968; Küpper-Tetzel et al., 2014).

Only a handful of classroom or classroom-like studies have examined the spacing effect with complex study materials as stimuli. For example, Yazdani and Zebrowski (2006) tested whether the scheduling of plane geometry homework (defined as either massed daily drilling after each covered topic or spaced homework over an extended period of time) would result in improved test scores six weeks later. The study's significant findings supported a shift towards a “non-drilling” method of instruction, strongly supporting spaced instructional design. Bird (2010) examined the ability of undergraduate learners of English to detect and correct verb morphology over a 14-week semester, with inter-study intervals of 3 and 14 days, and retention intervals of 7 and 60 days. After a 60-day retention interval, students benefited from the spaced (14 days) schedule of learning.

However, certain limitations of these studies should be mentioned. Bird's (2010) study, while having notable ecological validity, utilized five study sessions throughout the entire semester and engaged 5 h of class time, where students practiced the task of identifying mistakes in simple past/present perfect/past perfect sentences. While definitely a complex task and a well-designed experiment, it seems unlikely that this amount of class time would be spent on practice of the same material outside of this experiment. Also, this study had five study sessions, making it difficult to compare to the rest of the spacing effect literature, which typically employs two (or rarely three) study sessions. The same can be said for Yazdani and Zebrowski's (2006) study, which also had superior ecological validity and was conducted in a real high school classroom with actual school curriculum, but with less than ideal experimental control (e.g., the spacing manipulation had seven unequally spaced study sessions with an unequal amount of review completed at each session).

These studies highlight the difficulties researchers might encounter when translating psychological phenomena into educational settings. Time and curriculum constraints as well as practicality and efficacy of the way in which spacing can be implemented in the classroom often present challenges. In the current study, we propose a hybrid between a laboratory and a classroom setting as a step towards addressing some of these challenges.

1.4. Operational definition of higher-level learning

A major goal of the current study was to examine whether spaced review could improve higher-level learning; therefore, it was critical that we operationally defined what this term meant in the context of our study. We looked at the literature on critical thinking (e.g., Case, 2009; Ennis, 1987; Halpern, 2003; Kuhn, 1999; McPeck, 1981) to set operational definitions for “simple” and “complex” questions in the current study. After much consideration, we decided to use the Bloom's taxonomy of educational objectives framework (Bloom, 1956; Krathwohl, 2002) that is commonly used as the basis of assessment of student achievement in Canada (e.g., Ministry of Education, 2008). It consists of six hierarchical learning categories: Knowledge (recalling a fact), Comprehension (understanding meaning of a concept), Application (applying a concept to a new problem), Analysis (separating a concept into component parts), Synthesis (creating a new meaning or structure), and Evaluation (making judgements about the value of ideas or arguments). Our study required both simple and complex learning that could be objectively evaluated. We defined simple questions as those that assessed Bloom's Knowledge level. These were factual questions that only required students to retrieve information; they are referred to hereafter as factual. We defined higher-level questions as those questions that assessed Bloom's Application level. These were applied questions that required students both to retrieve *and* manipulate information in some way to solve a novel problem; they are referred to hereafter as higher-level. Sample factual and higher-level test questions can be found in Appendix A.

1.5. Current study

The current study incorporated a spacing manipulation into a simulated classroom environment using higher-level curriculum materials, educationally relevant inter-study and retention intervals, and time-efficient review methodology. We created a natural science lecture about meteorology and an online review with related test questions at two levels of complexity. We tested our spacing manipulation at inter-study intervals of 1 day and 8 days and with a retention interval of 35 days (chosen based on Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). Based on the existing published literature, we hypothesized that the spacing effect would extend to an undergraduate lecture-based classroom in that:

Students in the spaced condition would perform better on factual level questions (*Hypothesis 1*);

Students in the spaced condition would perform better on higher-level questions (*Hypothesis 2*).

2. Method

2.1. Participants

The participants were 175 undergraduate students enrolled in a first year psychology class. Data from three participants were removed because they grossly violated the study protocol (one student left for a 25 min bathroom break, thus missing 60% of the experiment; one student slept throughout the entire lecture; one student studied for a different class throughout the entire lecture). Data from another three participants were removed because of reported prior high familiarity with the lecture content. Therefore, data were analyzed for a total of 169 students (122 female) with a mean age of 22 years ($SD = 7$). Participants were awarded course credit for participation. We chose our sample size a priori to have 95% power at $d = 0.6$, our estimate of the effect size for our paradigm and sample.

2.2. Design

Although it was conducted in a classroom setting, this study conformed to the design of a typical spacing effect study. It consisted of an initial study session ("SS1", completed in a classroom, in the form of a lecture and an initial test), a review study session ("SS2", completed online either 1 day or 8 days after the initial study session), and a final test session (completed in the classroom 35 days after each review session). Students were randomly assigned to either the 1-day or 8-day spacing conditions, and students' performance on a short-answer test at the final test session acted as the dependent variable. In other words, in line with Hypotheses 1 & 2, the independent variables were: Spacing Condition (1-day vs. 8-day; manipulated between-subjects), and Complexity Level (factual versus higher-level; manipulated within-subjects). As an additional variable of interest and a manipulation check, we also tested whether a simple online review was a sufficient manipulation to produce a spacing effect benefit. This effect was examined by having participants review only half of the lecture material during the online review (SS2) and comparing students' performance on reviewed vs. not-reviewed questions at final test. In other words, we tested Effectiveness of Review (reviewed vs. not reviewed items; manipulated within-subjects) for both Complexity Levels (factual vs. higher level; manipulated within-subjects).

2.3. Materials

2.3.1. SS1: Lecture

For the initial learning session, a live, interactive, university-style lecture was created with the help of a meteorology textbook for a natural science course (Ahrens, 1999). Meteorology was chosen as a topic because it was likely to be novel, interesting, and attentionally engaging to introductory psychology students (these assumptions were verified by student self-report measures), and it was intricate enough that we could create questions assessing the two complexity levels. The 45-min lecture contained four units: clouds (formation and classification), wind (pressure streams and the Coriolis effect), rain (classification), and thunderstorms (stages of development; thunder and lightning). It was presented by a graduate student with teaching experience, was scripted and highly structured, and yet was designed not to overwhelm students with information. Breaks were provided throughout the lecture in the form of class participation and short videos (approximately 1–2 min). Class participation was limited to the instructor asking the class to raise their hands and vote on answers to various questions (e.g., a picture of a cloud was shown on the screen and the instructor asked, "Raise your hand if you think this is a cumulus cloud? A cirrus cloud? A nimbus cloud?"). Every effort was made to ensure that the lecture and the class environment resembled a typical undergraduate classroom as closely as possible. Apart from the initial 5 min of the study, where informed consent was signed and the procedure was explained, the lecture (in the authors' opinion) did not substantially differ from a regular class (see Section 4.2 for a detailed discussion on the similarities and differences between our lecture and a typical classroom environment).

2.3.2. SS1: Initial test

To gage whether students learned the content of the lecture by the end of the initial study session, a paper-and-pencil test was created with 40 short answer questions (20 factual and 20 higher-level).

2.3.3. SS2: Online review

The online review session was created using SurveyMonkey.com. The format of the review was as follows: (1) a short answer question (e.g., Q#1); (2) corrective feedback (re: Q#1); (3) a multiple-choice question mimicking the short-answer question (i.e., the same Q#1 question but with 4 choice options). Students were required to correctly answer the multiple-choice question in order to move on to the next question (i.e., Q#2; and so on until the end of the review). This format accomplished two purposes: First, it provided an opportunity for effortful retrieval (short answer question); second, it ensured that students actually read the corrective feedback (they could not move on until they provided correct answer to the multiple choice question). In accordance with principles of testing and feedback effects (e.g., Carpenter et al. 2009; Carpenter, Pashler, Wixted, & Vul, 2008; Pashler, Cepeda, Wixted, & Rohrer, 2005), structuring the online review in this way was an optimal means of re-learning in the context of this study.

The online review consisted of 20 short answer questions (10 factual and 10 higher-level) and their 20 complementary multiple-choice questions. Factual questions were repeated in exactly the same way as they were seen on the initial test. Higher-level questions were re-phrased to ensure that participants could not answer them correctly simply by recalling their answers from the previous test (Appendix A). Question re-phrases were piloted prior to the study to ensure that question difficulty did not change as a result of being re-phrased. All questions in the online review were presented in random order. It is important to note that students were given only half of the learned material to review (20 out of 40 questions from the initial test). This was our manipulation check to confirm the validity and effectiveness of the online review (discussed again in Section 3.1). Reviewed versus not reviewed questions were counterbalanced across participants.

2.3.4. Final test

At the final test session, students were given a paper-and-pencil, 40 question short answer test (20 factual and 20 higher-level) that encompassed all of the learned material. Factual questions were repeated in exactly the same way as on the initial test and online review. Higher-level questions were re-phrased for a second time.

2.4. Procedure

There were approximately 25 students per class. Students arrived to the classroom and took their seats. They read a general half-page overview of the lecture (to simulate lecture preparation) and were provided with a printed copy of the lecture slides (with ruled space to take notes beside each slide). Students were made aware of the initial test. Next, students listened to the lecture. At the end of the lecture, students were given the initial test and a demographics questionnaire. Students took, on average, 40 min to complete the test (although no time limit was enforced). The following day, students received an email asking them either to complete the online review (1-day condition) or to check their email again 7 days later to complete the online review at that time (8-day condition). Thirty-five days after completing their online review, students came back to the classroom and wrote the final test (again, no time limit was enforced).

3. Results

We have fully reported all data and analyses relevant to our hypotheses and all data that might affect interpretation of results. The dependent variable of interest was performance on the final test. Analyses included only the 20 questions that a given participant completed in all three sessions. As a reminder, the independent variables of interest were: Spacing Condition, manipulated between-subjects (1-day vs. 8-day), and Complexity Level, manipulated within-subjects (factual vs. higher-level). By chance, 76 participants ended up in the 1-day condition and 93 participants in the 8-day condition.

3.1. Manipulation check

In order for the spacing effect to take place, to-be-learned items need to be learned twice: once during the initial study session and once during the review study session. Initial test performance provided us with evidence of learning during SS1, and due to the structure of the SS2 online review session (i.e., participants could not move forward until they answered a question correctly), we assumed that all items were properly reviewed. Furthermore, we incorporated a test to check the effectiveness of the online review. Recall that the initial test and the final test contained 40 questions; however, the online review contained only 20 questions (i.e., 20 questions remained unreviewed). Because of this manipulation, comparisons could be made for items that were reviewed ($n = 20$) versus not reviewed ($n = 20$) during SS2. A 2 (Review: reviewed vs. not reviewed) $\times$ 2 (Complexity Level: factual vs. higher-level) repeated measures ANOVA confirmed that final test performance of reviewed questions was superior to final test performance of not reviewed questions, $F(1,168) = 181.34$, $p < .001$, $\eta^2 = .14$ (Table 1). There was a main effect of Complexity Level, $F(1,168) = 93.03$, $p < .001$, $\eta^2 = .22$, and no significant interaction, showing that review benefited both types of questions. Interestingly, we found that when Spacing Condition (1-day vs. 8-day) was entered as a between-subjects factor into the above 2 $\times$ 2 repeated measures ANOVA analysis, the Review $\times$ Spacing Condition interaction was not significant, $F(1,167) = .044$, $p = .834$. This was a surprising finding, as we expected that the spacing effect would only apply to questions that were reviewed during SS2. As such, additional analyses (contrasts) were conducted that revealed a spacing effect benefit for all final test questions, regardless of whether or not the questions were reviewed (see Table 2 and Section 4.2.1 for further discussion).

3.2. Spacing effect analyses

Two research assistants who were blind to participants' spacing condition graded each initial test and each final test. Inter-rater reliability was derived using Krippendorff's alpha (Hayes & Krippendorf, 2007). Reliability was very high, both for the initial test ($\alpha = 0.98$) and the final test ($\alpha = 0.99$).

To ensure no initial differences in performance between the 1-day and 8-day conditions, two *t*-tests were conducted on the initial test data. Results confirmed that the 1-day participants were not significantly different from the 8-day participants at the outset of the study, neither for factual, $t(167) = 0.16$, $p = .87$, $d = 0.03$ nor higher-level questions $t(167) = 1.12$, $p = .27$, $d = 0.17$.

Although the differences in initial test scores between spacing groups were not statistically significantly different, there was still a difference that slightly favored the 8-day condition. Therefore, to be conservative, we included initial test performance as a covariate in the final test analyses to control for any individual differences in learning ability. We conducted two one-way ANCOVAs, one for each level of complexity, because we had no a priori predictions of a Complexity Level $\times$ Spacing Condition interaction. In addition, a main effect of Complexity Level, if present, would not have a meaningful theoretical interpretation in the context of this study. Pertaining to Hypothesis 1, the one-way ANCOVA for factual questions revealed that there was a significant effect of spacing, $F(1,166) = 5.89$, $p = 0.02$, $\eta^2 = 0.026$. Participants in the 8-day condition ($M = 0.54$, $SD = 0.21$) performed better on the final test than participants in the 1-day condition ($M = 0.47$, $SD = 0.21$). In line with Hypothesis 2, the one-way ANCOVA for higher-level questions revealed that there was a significant effect of spacing, $F(1,166) = 4.71$, $p = 0.03$, $\eta^2 = 0.018$. Participants in the 8-day condition ($M = 0.42$, $SD = 0.18$) performed better on the final test than participants in the 1-day condition ($M = 0.35$, $SD = 0.20$). For a more conservative spacing effect analysis, see Appendix B.

4. Discussion

The results of this study provide evidence that spaced learning can be successfully implemented in a simulated classroom setting using realistic curriculum materials and educationally relevant

Table 1
Final test performance for questions that were reviewed vs. not reviewed in the online review.

Question level	Reviewed at SS2 *M* (*SD*)	Not reviewed at SS2 *M* (*SD*)
Factual	0.50 (0.21)	0.37 (0.20)
Higher-level	0.39 (0.19)	0.28 (0.17)

Table 2
Final test performance for questions that were reviewed vs. not reviewed in the online review as a factor of spacing condition.

Question type	Spacing condition	Reviewed M (SD)	Not reviewed M (SD)	t	p
Factual	1-day	0.47 (0.21)	0.33 (0.21)	7.36	<.001
	8-day	0.54 (0.21)	0.40 (0.19)	7.13	<.001
Higher-Level	1-day	0.36 (0.20)	0.25 (0.16)	6.82	<.001
	8-day	0.43 (0.18)	0.32 (0.17)	6.25	<.001

Note. $p < .0125$ was chosen as the level of significance for all contrast tests to correct for multiple comparisons.

inter-study and retention intervals. Participants who reviewed the lecture materials a week after initial learning performed better by one half-letter grade on a final test 35 days later when compared to participants who reviewed the lecture materials only one day after initial learning (7% for absolute test score in both question difficulty conditions, which is a 15% and 20% improvement for the 8-day condition over the 1-day condition baseline for factual and higher-level questions, respectively). The review was exactly the same for each participant; it was merely the manipulation of time – when the review was completed – that had an influence on long-term retention. In the same vein, this study highlights the importance of reviewing study materials via testing in general (see Table 1) and provides an uncomplicated method of incorporating review into teaching practice using a simple yet effective online exercise. In fact, given the integrated nature of the lecture, the questions that were not explicitly reviewed still received spacing effect benefits (see Table 2 and Section 4.2.1 for further discussion). We believe that a similar exercise (e.g., as a homework assignment) could be used to incorporate the benefits of spacing, testing, and feedback into real classroom practice.

4.1. Theoretical considerations

Currently, there is no unified spacing effect theory (Delaney, Verkoeijen, & Spirgel, 2010); however, contextual variability and study-phase retrieval theories are among the most prominent. Contextual variability theory (e.g., Glenberg, 1979) states that an item is stored in memory along with the specific context it was learned in (e.g., emotional state of the learner, physical environment, surrounding or related items to the to-be-learned item, etc.) and that context changes over time. The greater the number of unique contexts associated with an item, the greater the number of memory traces created for this item, and the more likely it will be successfully retrieved in the future. It is argued that spaced items, especially those with longer inter-study intervals, produce more variable contextual cues, thereby increasing the chance that at least some of them will be present at the time of the final test to serve as retrieval cues.

Study-phase retrieval theory (e.g., Thios & D'Agostino, 1976) states that re-learning of an item (i.e., the second presentation of an item) will only be successful if the first presentation is retrieved from memory and updated. For items that are retrieved immediately after initial learning, this reconstruction process will be relatively easy (having just seen the item). For items that are retrieved some time after initial learning, the reconstruction process will be more effortful, thereby more actively strengthening the memory representation of the item.

It is possible that in order to see similar effect sizes for higher-level and factual level questions, more than two study sessions are required (e.g., Bird, 2010). Even though we presented two higher-level practice questions on the same topic (during SS1 and SS2), the questions themselves were different (re-phrased), and we made the assumption that participants grasped that those two questions related to the same concept. For example, one higher-level question (Question 1 in Appendix A) required participants to recall a formula and apply it to two practice questions. What is being practiced is the process of applying the formula, not retrieving a single piece of information from memory, as is the case for factual level questions. This process could potentially be made up of a more advanced network of memory traces than the network required to retrieve a simple fact, and it is possible that this advanced network requires more than two practice sessions to build strong enough connections to exhibit a spacing effect of the same magnitude as that of the factual level questions. The question of why spacing effect size decreases as the complexity of a task increases (Donovan & Radosevich, 1999) is open to interpretation. Future research should look into exploring a potential interaction between the number of study sessions and spacing effect size as a function of question complexity.

4.2. From the laboratory to the classroom: Similarities and differences

One primary purpose of this study was to release some of the experimental control present in laboratory studies and establish whether the spacing effect is robust enough to be seen in a simulated classroom environment (i.e., a hybrid lab-classroom environment). Given this purpose, it is important to outline the major similarities and differences between this study and actual real-world classrooms.

4.2.1. Similarities

One of this study's unique contributions to the literature is that it combined a number of important classroom-like qualities while still maintaining a degree of experimental control. This study investigated the spacing effect in the context of group administration, in a classroom setting with a live lecturer, using real-world natural science curriculum with varying degrees of material complexity, and with a realistic method of implementing the spacing effect.

Most importantly, and most similar to real-world classrooms, presentation of information was done within a context rather than as a meaningless set of facts or word pairs. Presenting information within a context allows students to connect new information with pre-existing schemas. Because we had no way of predicting whether or how they incorporated presented material with their existing knowledge, the fact that spaced items still showed a memory advantage is especially noteworthy. (Although we controlled for extreme familiarity with the lecture material by asking participants to rate their familiarity with the lecture topic, the topic was "Everyday Weather Phenomena" and therefore we still assume at least some prior knowledge among all participants).

Additionally, it was surprising to find that questions not explicitly reviewed during SS2 also showed a spacing effect benefit (Table 2; For similar findings in the testing effect literature see, e.g., Chan, McDermott, & Roediger, 2006 and discussion in Little, Storm, & Bjork, 2011). While somewhat counterintuitive (why should questions that were not included in the online review show any spacing benefits?), this result highlights the integrative and holistic nature of the learning process. We believe that this result may have occurred due to the nature of our stimuli and test questions. Students' learning occurred in the context of a coherent lecture rather than studying isolated facts or practicing a single "complex" task. Also, the questions that were *included* in the review versus *not included* in the review were similar in nature and concerned the same topics. For example, the lecture material covered the characteristics of four cloud types; so, one reviewed factual question asked students "Which type of cloud is the *highest* in the sky?"

while a *not*-reviewed question asked students "Which type of cloud is the *lowest* in the sky?" Thus, to answer the question "Which type of cloud is the *highest* in the sky?" a student might need to effortfully retrieve *all* of the cloud types in order to write down the one that is the highest in the sky. Thus, the information for the cloud type that is the lowest in the sky, though not directly reviewed through answering the question and receiving feedback, might still have been implicitly reviewed through effortful retrieval. Since this effortful retrieval was strategically timed, "not reviewed" questions also showed spacing effect benefits.

4.2.2. Differences

Although many efforts were made to mimic a regular classroom, some differences still existed between our study and real-world learning. One difference was the inclusion of the initial test to measure participants' initial learning (and, in our opinion, solidify participants' memory of the topic through effortful retrieval). Participants not only were exposed to the lecture as passive listeners, but also were forced to recall information (factual questions) and actively think about novel applications of the information (higher-level questions) shortly after learning it (i.e., at the end of the lecture). Most university lectures do not have a test at the end of the class. Having a test or another exercise to solidify memory at the end of class is something researchers and teachers should take into consideration when conducting studies in the classroom or teaching new material.

Providing a solidifying exercise after presenting new material that includes the element of effortful retrieval (e.g., a pop quiz, a group activity, or even assigning a one-paragraph essay at the end of class on the topic of "What are the 5 most important things you learned in class today?") can act as an investment into better long-term retention. The testing effect (e.g., Roediger & Karpicke, 2006a; 2006b) suggests that testing students on recently learned material as a means of review engages them in a more effortful and active re-encoding experience than passive strategies such as having them re-read or be re-taught the material. Our decision to include an initial *test* and an online review *test* was motivated by this substantial literature.

Another important difference was students' motivation. This study created a stress-free learning environment. Students knew that they would receive course credit regardless of their test scores, and that they were participating in a one-time lecture research study, not an actual course that would show up on a transcript. On one hand, this could have alleviated some learning and test associated anxiety, and on the other hand, it could have resulted in less attention being paid to the lecturer. We doubt that this was the case, though; when asked, "How engaging did you find the lecture/lecturer?" 76% of students gave ratings of 5 and above (out of 7), where 1 = "Not at all engaging," 4 = "Somewhat engaging," and 7 = "Very engaging"). Regardless, the learning situation and students' motivation in this particular study varied from a typical classroom.

Finally, lecture quality differed between the current study and a regular classroom. Because the lecture acted as the main stimulus in this study, a considerable amount of time and effort was put into its creation and pilot testing to ensure that concept explanations were clear, an appropriate amount of time was spent on each concept, and student attention was maintained during the lecture. It would be unreasonable to expect a similar time commitment from instructors for any single lecture.

4.3. Challenges and limitations

The number of extraneous variables present in a classroom creates a challenge for any experimental study. In this study, some measures were taken to control for extraneous variables during the lecture. A research assistant (posing as a student) sat at the back of the classroom during the lecture and recorded behaviors that could interfere with the results of the study, for example, if a student left the classroom for an extended bathroom break, used a cell phone for texting, came to the study late, etc. Minimally disruptive behaviors were recorded and these participants still completed the study. As mentioned in Section 2.1, three participants who grossly violated study protocol (e.g., studied for a different class during the lecture, slept, etc.) were given course credit but were asked not to participate in the review and final test sessions of the study. Even with a research assistant and lecturer monitoring classroom activity, variances on final test scores were large, with standard deviations hovering around 20%, thereby reducing effect sizes and highlighting the challenge of spacing effect classroom implementation.

Perhaps another limitation of this study was the range of questions that we defined as higher-level. Using Bloom's taxonomy as a guide, our definition of higher-level questions was anything that required students to recall a piece of information and manipulate it or apply it in some way. Some higher-level questions were word problems (e.g., making an informed choice based on previously learned facts), some were calculation-based (e.g., applying a mathematical formula to solve a novel word problem), and others were induction-based (e.g., classifying an image of a cloud as one of four types). One might argue that different mechanisms are involved in solving these three types of questions and therefore they should be treated as different levels of higher-level thinking.

Finally, as discussed in Section 4.1, we cannot state with certainty whether the smaller spacing effect size for the higher-level questions was due to the questions' inherent complexity or to the fact that these questions were re-phrased at each session. Future studies exploring this issue should include a collection of "control" questions, where in addition to "pure" factual questions, half of the higher-level questions are re-phrased and half are not, thus having three levels of difficulty instead of two. A study designed in such a manner would provide a clearer answer to this question.

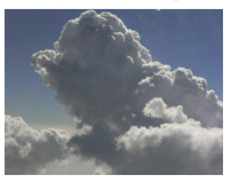
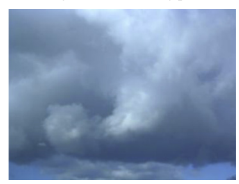
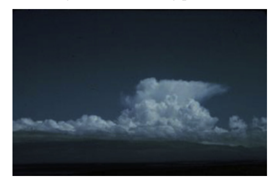
4.4. Conclusion

The current study contributes to the spacing effect literature by examining the phenomenon in a simulated university classroom, using natural science curriculum material and educationally relevant inter-study (1 day vs. 8 days) and retention (5 weeks) intervals. It provides evidence that the spacing effect is a robust phenomenon that can survive the added noise associated with a classroom environment and more complicated material. Simply by adjusting the timing of an online review, students in the optimally spaced condition improved their performance by one half-letter grade (7% absolute performance improvement). Instructors might take note of the simple strategies discussed in this study – especially providing well spaced out opportunities for effortful retrieval – and integrate them into their existing teaching practices.

Acknowledgments

We thank a number of dedicated volunteers and research assistants for their help with data collection and scoring: Adam Mishan, Alisa Rashkovan, Tammy Sikakane, and Lina Vishnevsky. We thank Irene Backhouse for assisting with classroom bookings. This work was funded by a Teaching-Learning Development Grant from the York University Faculty Association and by the York University Faculty of Health.

Appendix A. Sample test questions.

	Initial test (SS1)	Review (SS2)	Final test
Factual questions			
1.	What is a supercell?	What is a supercell?	What is a supercell?
2.	Warm air is associated with what type of air pressure?	Warm air is associated with what type of air pressure?	Warm air is associated with what type of air pressure?
3.	Evaporating streaks of precipitation are classified as what type of rain?	Evaporating streaks of precipitation are classified as what type of rain?	Evaporating streaks of precipitation are classified as what type of rain?
Higher-level questions			
1.	Mary hears sees lightning and hears thunder 12 s later. How far away is the storm from Mary?	Judy sees lightning and hears thunder 21 s later. How far away is the storm from Judy?	Jack sees lightning and a minute later he hears thunder. How far away is the storm from Jack?
2.	In which city would you predict a greater deflection of air molecules due to the Coriolis effect – Iqaluit, Nunavut, or London, Ontario? Why?	In which city would you predict less deflection of air molecules due to the Coriolis effect – Inuvik, Northwest Territories, or Toronto, Ontario? Why?	In which city would you predict a greater deflection of air molecules due to the Coriolis effect – Vancouver, B.C., or Yellowknife, N.W.T.? Why?
3.	Identify this cloud type.	Identify this cloud type.	Identify this cloud type.
			

Appendix B. Alternative spacing effect analyses.

In order for the spacing effect to emerge, to-be-learned items must be learned twice: once during the initial study session and once during the review study session. If for any reason learning does not occur during both SS1 and SS2, then the learning of that item has not been spaced. In particular, if a participant does not master a concept at initial learning, any review of this concept will not be review; it will be as if the participant is learning the concept for the first time at the review session. To this end, we ran an alternative spacing effect analysis to ensure that results were real and not due to noise. In this analysis we only counted an item as correct on the final test if a participant originally got the item correct on the initial test, thus satisfying the first assumption. When the final test data were analyzed in that manner, as pertaining to Hypothesis 1, the one-way ANCOVA for factual questions revealed a significant effect of spacing, $F(1,166) = 6.6$, $p = 0.01$, $\eta^2 = 0.016$. Participants in the 8-day condition ($M = 0.42$, $SD = 0.23$) performed better on the final test than participants in the 1-day condition ($M = 0.36$, $SD = 0.22$). As pertaining to Hypothesis 2, the one-way ANCOVA for higher-level questions revealed that there was a marginally significant effect of spacing, $F(1,166) = 3.6$, $p = 0.06$, $\eta^2 = 0.009$. Participants in the 8-day condition ($M = 0.31$, $SD = 0.18$) performed better on the final test than participants in the 1-day condition ($M = 0.25$, $SD = 0.18$). If, however, similar to lab studies (that use criterion learning), we examine the data only from participants who were paying most attention to the lecture (defined as students who scored 70% and above on the initial test, $n = 50$), the effect size increases approximately 6-fold from the complete data set for higher-level questions ($\eta^2 = .05$, small effect size) and 8-fold for the factual level questions ($\eta^2 = .13$, medium effect size).

References

Ahrens, D. C. (1999). *Meteorology today: An introduction to weather, climate and the environment*. California: Thompson Higher Education.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21*, 627–635. http://dx.doi.org/10.1080/00140137808931764.

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

Balch, W. R. (2006). Encouraging distributed study: a classroom experiment on the spacing effect. *Teaching of Psychology, 33*, 249–252.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31*, 635–650. http://dx.doi.org/10.1017/S0142716410000172.

Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York: David McKay Co, Inc.

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

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619–636. http://dx.doi.org/10.1002/acp.1101.

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. http://dx.doi.org/10.1002/acp.1507.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of testing on learning and forgetting. *Memory & Cognition, 36*, 438–448. http://dx.doi.org/10.3758/MC.36.2.438.

Case, R. (2009). Teaching and assessing the "tools" for thinking. In J. Sobocan, & L. Groarke (Eds.), *Critical thinking education and assessment: Can higher order thinking be tested?* (pp. 197–214). London, Ontario: The Althouse Press.

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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1111/j.14679280.2008.02209.x.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval induced facilitation: initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553–571. http://dx.doi.org/10.1037/0096-3445.135.4.553.

Delaney, P., Verkoeijen, P., & Spirgel, A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. *The Psychology of Learning and Motivation: Advances in Research and Theory, 53*, 63–147. http://dx.doi.org/10.1016/S0079-7421(10)53003-2.

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

DeRemer, P., & D'Agostino, P. R. (1974). Locus of distributed lag effect in free recall. *Journal of Verbal Learning and Verbal Behavior, 13*, 167–171. http://dx.doi.org/10.1016/S0022-5371(74)80041-1.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect. *Journal of Applied Psychology, 84*, 795–805. http://dx.doi.org/10.1037/0021-9010.84.5.795.

Dunlosky, K., & Rawson, A. (2012). Despite their promise, there's still a lot to learn about techniques that support durable learning. *Journal of Applied Research in Memory and Cognition, 1*, 254–256. http://dx.doi.org/10.1016/j.jarmac.2012.10.003.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). New York: W. H. Freeman and Company.

Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology, 59*, 290–296. http://dx.doi.org/10.1037/h0020055.

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

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

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology, 28*, 266–273. http://dx.doi.org/10.1002/acp.2997.

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

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

Hayes, A. F., & Krippendorf, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89. http://dx.doi.org/10.1080/19312450709336664.

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

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30*, 138–149. http://dx.doi.org/10.1086/374692.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*, 97–103. http://dx.doi.org/10.1002/acp.1801.

Kiepert, M. H. (2009). *An examination of repetition and the spacing effect in the classroom: A self-report survey of teachers* (Unpublished doctoral dissertation). Philadelphia, Pennsylvania: Temple University.

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

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592. http://dx.doi.org/10.1111/j.1467-9280.2008.02127.x.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice, 41*, 212–218. http://dx.doi.org/10.1207/s15430421tip4104_2.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*. http://dx.doi.org/10.3102/0013189X028002016, 12–26, 46.

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: enhancing students' memory for vocabulary. *Instructional Science, 42*, 373–388. http://dx.doi.org/10.1007/s11251-013-9285-2.

Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory, 19*, 346–359. http://dx.doi.org/10.1080/09658211.2011.569725.

Mackay, S., Morgan, P., Datta, V., Chang, A., & Darzi, A. (2002). Practice distribution in procedural skills training: a randomized controlled trial. *Surgical Endoscopy, 16*, 957–961. http://dx.doi.org/10.1007/s00464-001-9132-4.

McPeck, J. E. (1981). *Critical thinking and education*. New York: John Wiley.

Ministry of Education. (2008). *The Ontario curriculum, grades 11 and 12 science* (pp. 24–25). Retrieved from http://www.edu.gov.on.ca/eng/curriculum/secondary/science.html.

Moss, V. D. (1995). *The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student* (Unpublished doctoral dissertation). Logan, Utah: Utah State University.

Moulton, C. A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: what kind of practice makes perfect? *Annals of Surgery, 244*, 400–409. http://dx.doi.org/10.1097/01.sla.0000234808.85789.6a.

Panchuk, D., Spittle, M., Spittle, S., & Johnston, N. (2013). The effect of practice distribution and experience on the performance and retention of a discrete sport skill. *Perceptual & Motor Skills, 116*, 1–11. http://dx.doi.org/10.2466/23.30.PMS.116.3.750-760.

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. http://dx.doi.org/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 and Review, 14*, 187–193. http://dx.doi.org/10.3758/BF03194050.

Pyle, W. H. (1915). Concentrated versus distributed practice. *Journal of Educational Psychology, 5*, 247–258. http://dx.doi.org/10.1037/h0074998.

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

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55*, 297–308. http://dx.doi.org/10.1037/h0040734.

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

Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x.

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40*, 4–17.

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

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

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

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

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

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*, 341–358. http://dx.doi.org/10.1207/s1532690xci0103_4.

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

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

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Ozsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology, 22*, 685–695. http://dx.doi.org/10.1002/acp.1388.

Vlach, H. A., & Sandhofer, C. M. (2012). Distributed learning over time: the spacing effect in children's acquisition and generalization of science concepts. *Child Development, 83*, 1137–1144. http://dx.doi.org/10.1111/j.14678624.2012.01781.x.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*, 163–167. http://dx.doi.org/10.1016/j.cognition.2008.07.013.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. J. (2011). Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*, 750–763. http://dx.doi.org/10.3758/s13421-010-0063-y.

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

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: moderation by the difficulty of category discriminations. *Memory & Cognition, 41*, 16–27. http://dx.doi.org/10.3758/s13421-012-0238-9.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: application to exemplars presented as auditory and visual text. *Learning and Instruction, 22*(3), 215–221. http://dx.doi.org/10.1016/j.learninstruc.2011.11.002.