Educ Psychol Rev (2016) 28:853–873
DOI 10.1007/s10648-015-9349-8

INTERVENTION STUDY

Spaced Retrieval Practice Increases College Students' Short- and Long-Term Retention of Mathematics Knowledge

Robin F. Hopkins[1] **· Keith B. Lyle**[1] **· Jeff L. Hieb**[2] **·**
Patricia A. S. Ralston[2]

Published online: 12 December 2015

Abstract A major challenge college students face is retaining the knowledge they acquire in their classes, especially in cumulative disciplines such as engineering, where ultimate success depends on long-term retention of foundational content. Cognitive psychologists have recently recommended various techniques educators might use to increase retention. One technique (*spaced retrieval practice*) involves extending opportunities to retrieve course content beyond a customarily short temporal window following initial learning. Confirming the technique's utility requires demonstrating that it increases retention in real classroom settings, with commonly encountered educational content, and that gains endure into subsequent semesters. We manipulated spaced versus massed retrieval practice in a precalculus course for engineering students and followed a subset of students who proceeded into a calculus class the following semester. Spacing versus massing was manipulated within- and between-subjects. Within-subjects, students retained spaced content better than massed content in the precalculus course. Between-subjects, students for whom some retrieval practice was spaced, compared to those for whom all practice was massed, performed better on the final exam in the precalculus class and on exams in the calculus class. These findings suggest that spaced retrieval practice can have a meaningful, long-lasting impact on educational outcomes.

Keywords Memory · Spacing · Retrieval practice · Mathematics · Engineering

Results of this study were presented in part at the annual meetings of the Association for Psychological Science (New York, NY, May 2015) and the American Society for Engineering Education (Seattle, WA, July 2015).

✉ Keith B. Lyle
keith.lyle@louisville.edu

[1] Department of Psychological and Brain Sciences, University of Louisville, 2301 S. 3rd St., Louisville, KY 40292, USA

[2] Department of Engineering Fundamentals, University of Louisville, Louisville, KY, USA

Acquiring knowledge without retaining it is a fruitless venture, yet it characterizes the experience of many college students. College students learn large amounts of information in their classes but may quickly lose their ability to recall much of it (Bacon and Stewart 2006; Conway et al. 1991; Kamuche and Ledman 2005; Rawson et al. 2013). Although the consequences may sometimes be limited to disappointment and a sense of time wasted, they may at other times be more profound and even career path-altering. Success in some disciplines depends on students possessing a cumulative body of knowledge and is thwarted by poor retention of foundational content. Engineering is paradigmatic because success in engineering courses depends on prior mathematical knowledge. Pearson and Miller (2012) found that the two strongest predictors of completion of a baccalaureate in engineering were completion of a calculus course in high school and the number of calculus courses taken in college. Unfortunately, Pearson and Miller also found that nearly one third of students who enter an engineering program fail to complete it, and it has been argued that "the biggest factor contributing to the failure of engineering students is inadequate competence in mathematics" (Beanland 2010, p. 2). The implication is that inadequate retention of mathematics knowledge defeats the personal aspirations of many would-be engineers. Because excellence in science, technology, engineering, and mathematics is deemed critical for the USA' success in the twenty-first-century global economy (White House, n.d.), poor retention of mathematics may also undermine national priorities.

In recent years, concern over students' (collegiate and otherwise) poor retention of educational content has prompted cognitive psychologists to weigh in on the value of various techniques educators might use to foster greater retention. One oft-recommended technique is requiring students to retrieve learned information not only during initial acquisition and assessment periods, but later, after students have moved on to learning other material (e.g., Dunlosky et al. 2013; Pashler et al. 2007; Roediger and Pyc 2012). Carpenter (2014) gave this example:

> While learning about dependent-samples *t*-tests in a statistics course, students could receive a number of practice problems dealing with dependent-samples *t*-tests, in addition to some problems that cover the earlier-learned concepts of independent-samples *t*-tests and one-sample *t*-tests.

Advocates of this technique have pointed out that it is unlike what happens in many college classrooms, where material is often learned and tested in discrete, self-contained units or segments. Following acquisition of unit content, students are required to retrieve it multiple times—on homework, quizzes, tests, or all of these—but retrieval opportunities rarely extend beyond the end of the unit. In other words, retrieval practice is *massed* in a short temporal window following acquisition. In contrast, what has been recommended is more temporally distributed or *spaced* retrieval practice following acquisition.

We emphasize the post-acquisition nature of the recommended intervention because those who proffer the recommendation sometimes also recommend spacing study opportunities or early retrieval opportunities *during* initial learning. Although related, the two techniques are logically dissociable and post-acquisition spacing could be implemented regardless of whether initial acquisition activities were spaced or massed.

Enthusiasm for spaced retrieval practice originates in two intensively studied phenomena: the retrieval practice effect (also known as the testing effect) and the spacing effect. The retrieval practice effect refers to the robust finding that retrieving information from memory bolsters long-term retention more so than does restudying information (for a recent meta-analysis, see Rowland 2014). Scholarly interest in the phenomenon is long-standing (e.g., Gates 1917) and much recent research has focused on harnessing its power in classroom settings (e.g., Goosens et al. 2014; Lyle and Crawford 2011; McDaniel et al. 2013; McDermott et al. 2014). The spacing effect is the similarly robust (for a recent meta-analysis, see Cepeda et al. 2006) and long-established (Ebbinghaus 1964); phenomenon whereby increasing the temporal interval between learning events leads to enhanced retention (although the optimal interval beyond which further increases are not helpful depends on the retention interval; Cepeda et al. 2008). Classroom studies of this effect have also recently increased (e.g., Carpenter et al. 2009; Küpper-Tetzel et al. 2014; Sobel et al. 2011).

There is good empirical reason to believe that retrieval practice and spacing, in combination, could help students retain the information they labor to acquire. As referenced earlier, once individuals have grasped material sufficiently well to retrieve it from memory (suggesting at least some minimal degree of acquisition), additional retrieval increases the likelihood of long-term retention far more than additional restudy does (e.g., Karpicke 2009; Karpicke and Roediger 2008) and lengthening temporal intervals between retrieval opportunities (i.e., spacing) increases their beneficial impact (e.g., Cull 2005; Landauer and Eldridge 1967; Karpicke and Roediger 2007). While there is little question that spaced retrieval practice benefits human memory, efforts to establish its transformative power in authentic college settings remain preliminary. Pashler et al. (2007) described the evidence supporting the technique's usage in educational contexts as only moderate and, still more recently, Dunlosky et al. (2013) concluded that further systematic research was needed in genuine educational settings.

Although some studies have documented benefits of spaced retrieval practice in actual classroom settings or in the laboratory utilizing educationally relevant material, they cannot necessarily be taken to support the implementation of post-acquisition interventions in specific types of college courses, including math-intensive ones. Some studies manipulated spacing in the early stages of acquisition (e.g., Rickard et al. 2008) and others were conducted at the grade school or middle school level (e.g., Carpenter et al. 2009). While studies of a specific sub-type of spacing, known as interleaving, have repeatedly examined mathematics learning (e.g., Mayfield and Chase 2002; Taylor and Rohrer 2010), those studies implemented interleaving during the early stages of acquisition and furthermore, they did not concern retention of complex bodies of college-level mathematical knowledge (e.g., precalculus). Finally, we are not aware of any prior study that examined the effect of spaced retrieval practice on college students' performance in subsequent classes. This is an especially important limitation of the existing literature because, to be of value in cumulative disciplines such as engineering, spaced retrieval practice must increase retention not only within a given course but across semesters. Rawson et al. (2013) acknowledged the importance of examining effects of mnemonic interventions on long-term retention, but in Rawson et al.'s study of successive relearning, which involves spaced retrieval practice, the longest retention interval was only 24 days and did not extend into a subsequent semester. Promisingly, though, Rawson et al. found that successive relearning robustly increased retention of *Introductory Psychology* course content after 24 days, which minimally suggests that benefits of interventions which involve spaced retrieval practice are not fleeting.

Theoretically, the psychological mechanism by which spaced retrieval enhances long-term retention is the subject of ongoing debate. Multiple mechanisms have been proposed and the present study was not designed to arbitrate between them. However, one idea we see as potentially important in the context of classroom mathematics learning is study-phase retrieval, which has a long history of consideration (Thios and D'Agostino 1976) and remains of active interest (Cepeda et al. 2006; Delaney et al. 2010; Toppino and Gerbier 2014). As applied to spaced retrieval practice, study-phase retrieval is the idea that retrieving information (e.g., the steps in performing a mathematical operation) is more likely to involve actively retrieving a memory of the initial study experience (e.g., reading about the operation in a textbook) when the retrieval attempt occurs at a longer temporal interval following initial study. At a relatively short temporal interval, the study experience retains some of its initial activation and retrieval of that experience may be either unnecessary or relatively effortless. This makes intuitive sense in the context of classroom mathematics learning. Soon after learning how to perform a mathematical operation, the steps are fresh in students' minds and students may implement those steps with relatively little reflection. At a delay, when the operation has long been out of students' working memory, students will need to retrieve the instructional material they initially studied from episodic memory. This could have various consequences (not mutually exclusive) that could enhance retention. For example, the material could acquire new contextual associations and/or the need for active retrieval could be metacognitively interpreted as difficulty, triggering analytic processing of the material (e.g., comparing and contrasting it with other material; Alter et al. 2007).

Overview of the Present Study

In the present study, we manipulated spaced versus massed retrieval practice in an *Introductory Calculus for Engineers* course. This was a precalculus course created specifically to serve the needs of first-semester freshmen who were admissible to the engineering college but either had math ACT scores below that required for the first calculus course or otherwise felt themselves unprepared for the first calculus course. The content of Introductory Calculus for Engineers is foundational for subsequent engineering math courses and its retention is critical for long-term success in engineering. Massing or spacing of retrieval practice was implemented on weekly quizzes by varying the distribution of problems targeting particular learning objectives. Solving these problems required retrieving objective-critical information. The quizzing stage can be considered post-acquisition because students were required to demonstrate a minimal degree of mastery of each objective before taking quizzes.

Spaced versus massed retrieval practice was manipulated in a hybrid between- and within-subjects design. All students were taught in the same class by the same instructor, but were assigned to one of two groups that did not differ significantly on any of several potentially important preexisting characteristics. For students in a control group, retrieval practice for all learning objectives was massed, as it had been in previous iterations of the course. For students in an experimental group, retrieval practice was massed for some learning objectives but spaced for others (henceforth, massed versus spaced objectives, respectively). Assignment of objectives to spacing or massing was counterbalanced.

In both groups, students received immediate feedback on their quiz performance. Students were shown the questions they answered incorrectly and they could opt to view the correct answers, although they were not required to do so. Hence, the current study concerned effects of massed versus spaced retrieval practice with feedback.

Our experimental design allowed us to test the effect of spaced retrieval practice in multiple ways. Obviously, we could assess the intervention's effectiveness in within-subjects fashion by examining whether individuals in the experimental group retained spaced objectives better than massed ones. Also fairly obviously, we could compare across groups and see whether retention of spaced objectives in the experimental group was greater than retention in the control group, in which all objectives were massed.

Our primary reason for adopting a hybrid within- and between-subjects design, however, was that it afforded a third, less obvious comparison. Namely, we could compare retention of massed objectives in the experimental group to retention in the control group. We were interested in the possibility that the experimental group's retention of massed objectives would exceed the control group's. Given the interrelated nature of learning objectives in a precalculus course, it seemed possible that strengthening students' grasp on some objectives via spaced retrieval practice would incidentally help students master other, nonspaced objectives. For example, having a firm grip on polynomial long division could help students master oblique asymptotes. This is in line with recent theorizing that retrieval practice, being an effective means of increasing knowledge retention, also increases flexible usage of that knowledge, allowing people to transfer knowledge to novel settings and make novel connections (Carpenter 2012; Karpicke 2012). If spaced retrieval practice increases retention even more so than massed retrieval practice, then it might further assist people in learning related information. The notion that a mnemonic intervention might increase retention, not only of targeted information, but also of related information, is not without precedent. Prior studies have documented retrieval-induced facilitation whereby practicing retrieval of some information from an integrated body increases retention of the practiced information as well as related, nonpracticed information (Chan et al. 2006; Cranney et al. 2009; see also Rowland and DeLosh 2014).

For all of the comparisons described earlier, the criterial test of retention was the cumulative final exam in Introductory Calculus for Engineers, which was administered weeks to months (range=2.5 weeks to 3 months) after information was initially acquired. This is a long interval relative to that in many laboratory studies, and even in prior classroom studies (e.g., Rawson et al. 2013), but it is short relative to the multi-year engineering curriculum. While retaining course material until the final exam is important for success in the course itself, it would not be sufficient for success in the discipline of engineering. As part of the current study, we also sought to assess the effect of spaced retrieval practice on longer-term retention. Therefore, we followed a subset of students who went on to take *Engineering Analysis I* the next semester (i.e., the second semester of students' freshman year). Engineering Analysis I is the first calculus course for engineering students in which the fundamental concepts of differentiation and integration are developed and related via the fundamental theorem of calculus. Applications of these concepts are used to solve engineering problems, including ones involving motion, related rates, optimization, and moments and centers of mass. Mastering many (but not all) learning objectives in Engineering Analysis I requires retaining information acquired in Introductory Calculus for Engineers. We therefore examined three performance measures in Engineering Analysis I, described below.

Our first measure was the first unit exam in the class. The first unit in Engineering Analysis I primarily covered material from Introductory Calculus for Engineers. If students who had been in the experimental group retained material from the previous semester better than students who had been in the control group, then the former should score higher on the first unit exam in Engineering Analysis I. This exam was administered approximately 4 weeks after the final exam in Introductory Calculus for Engineers and we considered it a measure of intermediate retention. Second, to test the specificity of the effect of spaced retrieval practice, we examined performance on the third unit exam in Engineering Analysis I, which was the first in the course to exclusively cover material that did not directly build on any spaced learning objectives from Introductory Calculus for Engineers. We expected no difference on this exam between subjects who had been in the experimental group versus the control group. Third, we examined performance on the cumulative final exam. Because this exam covered many objectives whose mastery depended on knowledge retained from Introductory Calculus for Engineers, we expected higher performance on the exam by subjects who had been in the experimental group versus the control group. The final exam, which was administered 4.5 months after the final exam in Introductory Calculus for Engineers, served as our measure of long-term retention in this study.

Our study is not the first to ask whether spaced retrieval practice can help improve the classroom performance of engineering students. A recent study by Butler et al. (2014) included spacing of homework problems as one part of a three-part composite intervention aimed at enhancing learning in a math-intensive engineering course (*Signals and Systems*). The other two parts of the intervention were increasing the amount of retrieval practice (concomitant with its spaced temporal distribution) and increasing students' exposure to corrective feedback. The intervention was manipulated within-subjects, as in our experimental group, and at the level of particular topics, akin to our learning objectives. On two exams in the class, Butler et al. found that students performed better on questions covering topics to which the intervention had been applied. Both exams were non-cumulative and hence assessed relatively short-term retention.

Butler et al.'s study is extremely encouraging in showing that the application of principles from cognitive psychology can enhance performance in an engineering course, but it was not designed to isolate the effect of any one part of its three-part intervention. Theoretically, spaced retrieval practice itself could have had no effect on performance and the intervention's efficacy could have been driven entirely by one or both of the other parts. Hence, the study leaves open the question of whether spaced retrieval practice alone is sufficient to increase retention. Nor did the study address whether spaced retrieval practice can increase long-term retention and, in particular, retention into subsequent semesters. The present study was designed to address these important unanswered questions.

Method

This research was approved by the Institutional Review Board at our institution. An independent evaluator monitored the research to ensure that students assigned to the control group received fair treatment, despite having spacing withheld from their instructional plan in Introductory Calculus for Engineers.

Subjects

All subjects were enrolled in Introductory Calculus for Engineers in the Fall 2014 semester, with a subset also having been enrolled in Engineering Analysis I in the Spring 2015 semester. Prior to the start of the semester in which Introductory Calculus for Engineers was offered, subjects were randomly assigned to the experimental or control group by the first author, who had no contact with students at any point. Subsequent to this initial random assignment, the group assignment of a small number of subjects was changed by the first author to ensure that the groups were as comparable as possible on racial composition, gender composition, mean ACT Math score, and mean high school grade point average (GPA). Changes in group assignment were based solely on those four factors. Equating the groups on these factors was deemed important because any or all might plausibly be related to performance in college mathematics courses. ACT Math score is an especially important characteristic to equate, because it is known to predict performance in Engineering Analysis I (Hieb et al. 2015). Via this process, 58 students were assigned to each group but, during the course of the semester, 22 students withdrew from the class or otherwise stopped participating (a typical number for this class). Of the withdrawn or nonparticipating students, 14 had been assigned to the experimental group and eight to the control group, a nonsignificant difference, $\chi^2(1)=1.402$, $p=.24$. Data from these subjects were not included in any analyses. Table 1 shows the demographic and academic characteristics of those students who completed Introductory Calculus for Engineers. The groups did not differ significantly on any characteristic, all ps$>.90$.

Table 1 also shows the number and characteristics of students who proceeded to Engineering Analysis I after completing Introductory Calculus for Engineers. The proportion of students proceeding from each group did not differ significantly, $\chi^2(1)=0.012$, $p=.912$. The groups remained statistically indistinguishable on academic and demographic characteristics at the start of Engineering Analysis I, all ps$>.56$.

Course Format and Materials

Introductory Calculus for Engineers The course format followed the NCAT emporium model (http://www.thencat.org/R2R/AcadPrac/CM/MathEmpFAQ.htm). The course was divided into six units, each lasting approximately 2 weeks. In each unit, students read sections from an e-book (*Precalculus*: *A Right Triangle Approach* by Kirk Trigsted), completed practice problems, and tested themselves as part of structured sequences called study plans, and took two quizzes and a unit exam (see below for details on study plans, quizzes, and unit exams). There were no lectures. Students attended a single weekly class meeting in which the course instructor reviewed the schedule of upcoming assignments, discussed administrative issues, answered questions, and discussed the relevance of course topics to the field of engineering. Outside of these meetings, students were required to spend at least 150 min per week in a laboratory setting, working independently on study plans. The laboratory was staffed with teaching assistants who answered questions and helped students work toward mastering course objectives.

Table 1 Demographic and academic characteristics of subjects meeting various criteria

Criterion	Group	*N*	White (%)	Male (%)	Mean math ACT (SD)	Mean high school GPA (SD)
Completed ICE	Control	50	78.0	68.0	25.6 (1.12)	3.77 (0.66)
	Experimental	44	77.3	65.9	25.6 (1.37)	3.76 (0.44)
Took all ICE exams	Control	46	80.4	69.9	25.6 (1.13)	3.83 (0.65)
	Experimental	40	77.5	67.5	25.6 (1.14)	3.77 (0.44)
Proceeded to EAI	Control	37	78.4	64.9	24.5 (1.10)	3.89 (0.69)
	Experimental	33	72.7	69.7	25.5 (1.28)	3.80 (0.43)
Took all EAI exams	Control	29	72.4	62.1	25.3 (0.76)	3.95 (0.69)
	Experimental	25	76.0	64.0	25.3 (1.22)	3.78 (0.38)

ICE Introductory Calculus for Engineers, *EAI* Engineering Analysis I

Introductory Calculus for Engineers contains a total of 214 learning objectives identified prior to commencement of the present study. These objectives were distributed roughly equally across the six units. The third and fourth authors selected 12 objectives from each of the first four units (48 total) which they deemed especially critical for success in the course and foundational for the mathematics sequence in engineering. These were designated target objectives (see Table 2 for examples). The remaining objectives were designated nontarget. Nontarget objectives were presented in the course as usual but retrieval practice for them was not manipulated. All objectives from the fifth and sixth units were designated nontarget because they were presented too late in the semester to receive the spacing manipulation (see below).

Questions covering learning objectives were administered on study plans, quizzes, unit exams, and the cumulative final exam. We describe each of these in turn, but note first that all were delivered via MyMathLab®, an online media and homework system developed by textbook publisher Pearson. Regarding study plans, there were two such plans per unit. Study plans were collections of practice problems and a self-testing mechanism (called Quiz Me) covering the learning objectives presented within each unit's e-book reading. Learning objectives were covered sequentially within the study plan such that students first completed practice problems for a specific objective and then, when they felt they had mastered the objective, they performed the Quiz Me activity. Quiz Me presented students with three questions covering the just-practiced objective. Quiz Me questions were similar to practice problems. If a student scored above 70 % on the Quiz Me questions, then the objective was marked as mastered and the student could advance to the next objective in the plan. Partial credit was possible on Quiz Me questions. Consequently, reaching the 70 % criterion often, but not always, involved answering all three questions targeting a given objective correctly. Study plans for the first four units in the course contained practice problems and a Quiz Me activity for target and nontarget objectives. Study plans for the fifth and sixth units covered nontarget objectives only. Study plans were the same for all students, regardless of group assignment.

For each study plan, there was a corresponding quiz. Because there were two study plans per unit and there were six units, there were 12 quizzes in total. Quizzes were distinct from and in addition to Quiz Me activities. Three quiz questions assessing mastery of each target objective were drawn from quizzes and tests administered in previous iterations of Introductory Calculus for Engineers. Some adjustments to questions were made based on the third and fourth authors' expertise to ensure that all questions were of comparable difficulty. The adjustment process was also informed by objective data yielded by MyMathLab® in previous semesters. Quiz questions were similar to practice problems and Quiz Me questions from the study plan and questions covering the same target objective were highly similar to one another,

Table 2 Example target objectives in introductory calculus for engineers

Objective description
Using the order of operations to simplify numeric and algebraic expressions
Solving quadratic equations by factoring and the zero product property
Determining the domain of a function given the equation
Understanding the definition of a logarithmic function
Sketching graphs of the form $y=A\sin(Bx-C)+D$ and $y=A\cos(Bx-C)+D$
Given the graph of a function, find designated limits and state its value at specified points

often containing nearly identical wording and usually differing only in terms of the numbers that were part of the problem.

Our experimental manipulation was implemented within the quizzes. In the control group, each quiz administered in the first four units of the course contained three questions covering each of six target objectives, all of which were introduced to students in the corresponding study plan. In other words, questions covering each target objective were massed (see Table 3). Quizzes in the first four units also included up to eight additional questions covering nontarget objectives. Quizzes administered to control-group subjects in the fifth and sixth units contained questions covering nontarget objectives only.

Quizzes in the experimental group were identical to those in the control group except in the distribution of questions covering half ($n=3$) of the target objectives. For these objectives, quiz questions were spaced as described next and as depicted in Table 3. One question appeared on the quiz corresponding to the study plan in which the objective was initially presented. A second question appeared on the subsequent quiz. The next quiz in sequence did not contain any questions covering these objectives but the quiz thereafter contained the third and final question. We chose to skip one quiz and thereby delay presentation of the final question because we thought it might make the final retrieval more effortful. We were inspired by theorizing that retrieval-practice effects are greater when they require greater retrieval depth (Karpicke and Roediger 2007). Quiz questions for the remaining target objectives were massed, exactly as in the control group. Critically, assignment of objectives to the massed or spaced condition in the experimental group was counterbalanced. Note that, due to the spacing manipulation, quizzes administered to experimental group subjects in the fifth and sixth units, unlike those administered to control-group subjects, contained questions covering both nontarget and target objectives.

Each unit (except the sixth) concluded with administration of a unit exam. Exams for the first four units contained questions covering both target and nontarget objectives. The exams for the fifth unit covered nontarget objectives only. Unit exams were the same for all students, regardless of group assignment.

A cumulative final exam covered all target objectives along with a subset of nontarget objectives. Most target objectives were covered by a single question but some were covered by as many as three questions. When multiple questions covered the same objective, proportion correct for those questions was calculated to represent a student's mastery of the objective. Partial credit was possible when there were multiple steps involved in a question (e.g., finding the quotient and remainder in polynomial long division, or finding the value of the six trigonometric functions given a point on the unit circle). Where appropriate, un-simplified

Table 3 Representative depiction of the temporal distribution of quiz questions in Introductory Calculus for Engineers

Group	Objective and condition	Quiz			
		1	2	3	4
Experimental	A – Spaced	Question 1	Question 2	–	Question 3
	B – Massed	Questions 1, 2, and 3	–	–	–
Control	A – Massed	Questions 1, 2, and 3	–	–	–
	B – Massed	Questions 1, 2, and 3	–	–	–

A and B represent two distinct target learning objectives. Assignment of A and B to the Spaced or Massed condition was counterbalanced in the Experimental group. Only the first four quizzes in the course are depicted answers were scored as 95 % correct. The final exam was the same for all students, regardless of group assignment.

Engineering Analysis I Students attended three lectures each week and a review session prior to each unit exam. The class used a custom textbook; information about the text is available upon request from the authors. There were weekly homework assignments and quizzes administered via MyMathLab® but these were not the focus of any of our analyses. For our purposes, the critical materials in this class comprised three exams. All were administered via paper and pencil. Two were unit exams—the first and third out of a total of 13 administered in the semester. The first unit exam covered learning objectives presented in Introductory Calculus for Engineers. The third unit exam did not cover any of such objectives. We also analyzed performance on the cumulative final exam.

Procedure

Introductory Calculus for Engineers The syllabus for Introductory Calculus for Engineers informed students of the study and explained that they could opt not to have their performance data included in analyses by contacting their instructor. No students chose to have their data excluded. Students were informed that the study concerned "spaced retrieval" but no information regarding hypotheses was provided and students did not know to which group they had been assigned. All students, regardless of group assignment, attended the same class meetings, had the same lab attendance requirements, and were taught by the same instructor and teaching assistants.Students worked on study plans in lab and at other times and places of their choosing. Hence, completion of Quiz Me activities was unproctored.

Quizzes were administered once a week during a 48-h window that was the same for all students, regardless of group assignment. Critically, this means that students assigned to the control group had to answer all quiz questions covering a given target objective within a single 48-h window. In contrast, students assigned to the experimental group answered quiz questions for only half of all target objectives in this massed fashion. Quiz questions for other target objectives within the experimental group were administered in spaced fashion such that there was 1 week between administration of the first and second questions and 2 weeks between administration of the second and third questions. Before students could take any quiz, at least 80 % of the learning objectives in the student's corresponding study plan for that quiz had to be marked as mastered, as described under the "Course Format and Materials" section. This requirement was the same for all students, regardless of group assignment.

Quizzes were unproctored and students had 90 min to complete them. A browser lockout prevented students from accessing any resources on their computers during quizzes. Students were encouraged to take quizzes without referring to notes or the textbook but this was not monitored or enforced. Immediately upon completion of each quiz, students were re-presented with the list of quiz questions. Questions answered incorrectly (in whole or in part) were indicated by a red X. Students could click a link to view correct answers. Students were required to take each quiz twice during the quiz's allotted 48-h window. This requirement existed because students were instructed to use their first attempt to assess their understanding of the material and identify areas for improvement. Both quizzes covered the same set of objectives and had the same number of questions. Questions were algorithmically generated so they differed between the two instantiations of the quiz.

Five unit exams were administered in a proctored classroom setting at 2-week intervals with a 60-min time limit. There was no unit exam following the sixth unit. Rather, the cumulative final exam was administered 1 week after the closing date for the second quiz in the sixth unit. The final therefore occurred 1 week after the last quiz containing spaced questions over target objectives (in the experimental group) and 5 weeks after the last quiz containing massed questions over target objectives (in both the experimental and control group). The final was administered in a proctored classroom setting with a 2.5-h time limit. The time and place of unit exams and the final exam was the same for all students, regardless of group assignments.

Answers on all quizzes, unit exams, and the final exam were scored by MyMathLab®, precluding the possibility of more favorable scoring for objectives or students exposed to spacing.

Engineering Analysis I All exams were administered in a proctored setting. The time limit for exams was 1 h for unit exams and 2.5 h for the cumulative final exam. Graders were blind to students' group membership in Introductory Calculus for Engineers, precluding the possibility of more favorable scoring for students who had been in the experimental group.

Results

Introductory Calculus for Engineers

To assess retention of critical course content acquired throughout the semester, we analyzed performance on final exam questions that covered target objectives (ignoring questions which covered nontarget objectives). For each student, we calculated the proportion of questions answered correctly. In the experimental condition, proportion correct was calculated separately for questions covering objectives which were spaced versus massed. In the control condition, in which all objectives were massed, a single proportion correct score was calculated.

We restricted our analysis of final exam performance to those students who had taken all preceding unit exams throughout the semester. We did this to ensure that the experimental and control groups were equated on amount of retrieval practice, since unit exams required students to retrieve objective-critical information. This criterion resulted in the exclusion of four students from the experimental group and four students from the control group. Table 1 shows the academic and demographic characteristics for retained students. The groups did not differ significantly on any of these variables, all ps>.62, but we nonetheless utilized them as covariates in all analyses reported below.[1] Because there were few nonwhite students, race was dummy coded simply as White (1) or non-White (2). Gender was dummy coded as male (1) or female (2). Estimated marginal means are reported.

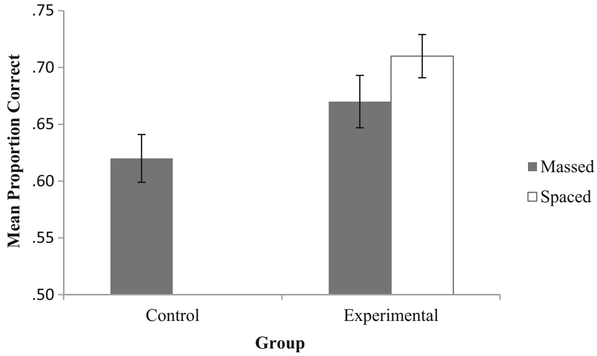
[1] With two exceptions, no covariate had a significant main or interactive effect on any performance measure in either Introductory Calculus for Engineers or Engineering Analysis I. One exception is a significant interaction between gender and objective type in the within-subjects analysis of final exam performance in Introductory Calculus for Engineers, $F(1, 35)=9.77$, $p=.004$, $\eta^2=.183$. The interaction arose because a significant difference in performance between spaced and massed objectives occurred for male but not female students. The other exception is that high school GPA was significantly associated with performance on the first unit exam in Engineering Analysis I, $F(1, 64)=4.63$, $p=.035$, $\eta^2=.063$. Both findings, while potentially interesting, may be spurious because, one, the number of female students was less than 20 ($N=13$) and hence too small to instill much confidence in the result (see Simmons, Nelson, and Simonsohn 2011), and, two, high school GPA was not significantly associated with any other performance measure.

Within-Subjects Analysis We first tested the effect of spaced retrieval practice within-subjects by comparing performance on spaced versus massed objectives within the experimental group. Proportions correct for the two objective types were submitted to a repeated-measures analysis of covariance (ANCOVA). Delaney and Maxwell (1981) showed that repeated-measures ANCOVA is vulnerable to a statistical artifact whereby the presence of covariates can cause the main effect of the repeated measure (in this case, objective type) to appear weaker than it actually is (for further discussion, see Thomas et al. 2009). Delaney and Maxwell provided a method of eliminating this artifact, which is to mean center the covariates. We applied this solution to the two covariates for which mean-centering was sensible: Math ACT score and high school GPA. The repeated-measures ANCOVA yielded a significant main effect of objective type, $F(1, 35)=7.96$, $p=.008$, $\eta^2=.150$. The rightmost pair of bars in Fig. 1 shows that, as predicted, proportion correct on spaced objectives ($M=.71$) was higher than on massed objectives ($M=.68$).

Between-Subjects Analyses We next compared performance on spaced objectives in the experimental group to performance in the all-massed control group (see Fig. 1). A between-subjects ANCOVA indicated that proportion correct on spaced objectives ($M=.71$) was significantly higher than on massed ones ($M=.63$), $F(1, 80)=9.01$, $p=.004$, $\eta^2=.099$ In a separate between-subjects ANCOVA, we compared proportion correct in the all-massed control group to proportion correct on massed objectives in the experiment group. Proportion correct was higher in the experimental group ($M=.68$) and this difference approached significance, $F(1, 80)=3.71$, $p=.058$, $\eta^2=.043$.

Although we expected to find these between-subjects differences, one might wonder whether the experimental group's advantage over the control group was due, not to spaced retrieval practice, but to some unanticipated difference between members of the two groups in terms of their preexisting aptitude for learning precalculus (despite their being equated on Math ACT scores and high school GPA). We therefore compared the groups' performance on the first unit exam in Introductory Calculus for Engineers, which was administered at a time when the experimental group's exposure to spaced retrieval practice was minimal. Performance in the experimental group ($M=.66$) was numerically better than in the control group ($M=.59$) but the difference did not approach significance, $F(1, 80)=1.53$, $p=.220$, $\eta^2=.018$. Readers may notice that, despite being nonsignificant, the mean performance difference is almost as large as between experimental-spaced and control-massed objectives on the final exam. Given that the difference on the final exam was highly significant, it is curious that the

Fig. 1 Mean proportion correct for spaced and massed objectives on the final exam in Introductory Calculus for Engineers. *Error bars* indicate ±1 SEM. Note that no objectives were spaced in the control group



difference on the first unit exam did not approach significance. The explanation appears to lie in the fact that performance on the first unit exam was more variable than on the final exam, making the mean a less informative measure of central tendency on the former than the latter. We therefore examined median scores, as well. On the first unit exam, the median was .68 in the experimental group, barely higher than the median of .66 in the control group. On the final exam, in contrast, the median for experimental-spaced objectives was .72 versus .63 for control-massed objectives.

As a final illustration of the value of spacing in Introductory Calculus for Engineers, we calculated subjects' proportion correct collapsed across spaced and massed objectives in the experiment group (M=.70) and compared it to proportion correct on target objectives in the control group (M=.63). A between-subjects ANCOVA yielded a significant effect of group, $F(1, 80)=6.58$, $p=.012$, $\eta^2=.074$. This indicates that mastery of target objectives was overall superior among students who received spaced retrieval practice versus those who did not.

Quiz Performance Although not a focus of our research, we also analyzed performance on the 12 quizzes. Our analysis was conducted on the same subset of students for whom we analyzed final exam performance. Students spent considerable amounts of time taking quizzes. The median completion time, depending on quiz, ranged from 24 min and 25 s to 55 min and 6 s. We submitted proportion correct on quizzes to a 2 (group: experimental or control) × 2 (quiz attempt: first or second) mixed-design MANCOVA in which the first factor was between-subjects and the second was within. The covariates were the same as in previous analyses. Occasionally, subjects did not attempt a quiz or submitted only one attempt instead of two. In these rare cases (a total of 33 missing attempts out of 2064, or 1.6 %), we replaced the missing score with the average proportion correct for all students who submitted the particular attempt. There were no significant effects, although the effect of group approached significance, $F(12, 69)=1.84$, $p=.059$. This trend was driven by the fact that, on quiz 8, subjects in the control group (M=.86) significantly outperformed subjects in the experimental group (M=.81), $F(1, 80)=7.52$, $p=.008$, $\eta^2=.080$. Despite this difference, the estimated marginal mean averaged across all quizzes (and collapsed across attempt) was exactly the same (M=.78) for both groups.

Engineering Analysis I

We submitted each DV to a separate between-subjects ANCOVA with the same covariates as in preceding analyses. Our analyses of the first and third unit exams were conducted on data from all students who advanced from Introductory Calculus for Engineers (see Table 1). Our analysis of final exam data from Engineering Analysis I included only those subjects who took all unit exams in the course, following the same reasoning put forth above regarding Introductory Calculus for Engineers. This resulted in the exclusion of five individuals from the control group and six from the experimental group. Table 1 shows the academic and demographic characteristics of retained students, which did not differ as a function of group, all ps>.29.

Figure 2 shows that our findings conformed to expectations. Students who had been in the experimental group in Introductory Calculus for Engineers scored higher on the first unit exam (M=.62) than students who had been in the control group (M=.56), although this difference

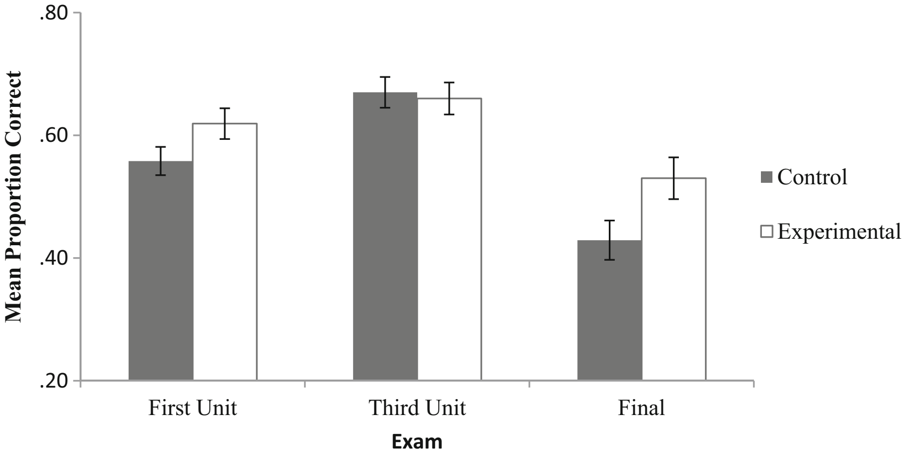


Fig. 2 Mean proportion correct on three exams in Engineering Analysis I for students formerly in the control and experimental groups in Introductory Calculus for Engineers. *Error bars* indicate ±1 SEM

only approached significance, $F(1, 64)=3.21$, $p=.078$, $\eta^2=.043$. In marked contrast, the groups performed nearly identically on the third unit exam, which covered material that was independent of the target objectives in Introductory Calculus for Engineers (*M*s=.66 and .67 for experimental and control subjects, respectively), $F(1, 64)=.073$, $p=.787$, $\eta^2=.001$. Performance on the cumulative final exam revealed a significant advantage for students who had been in the experimental group ($M=.53$) versus the control group ($M=.43$), $F(1, 48)=4.68$, $p=.035$, $\eta^2=.087$.

To assess the long-term impact of spaced retrieval practice in a manner that is especially meaningful to educators and others concerned with student success, we also looked at the proportion of students earning a satisfactory final letter grade in Engineering Analysis I as a function of group membership in Introductory Calculus for Engineers. Our definition of satisfactory (see also Paunesku et al. 2015) was a C or higher on an A–F scale on which it was also possible to receive plusses or minuses. Hence, a C− or lower was coded as unsatisfactory. Only students who completed all exams in Engineering Analysis I was included in this analysis. As shown in Table 4, only a minority of students who had been in the control group earned a satisfactory grade versus a majority of students in the experimental group. We analyzed these results in a multiple logistic regression. As predictors, we first entered the factors we have used as covariates in preceding analyses and then added group membership. The full model was not significant, $\chi^2(1)=1.81, p=.179$. Although group membership was not a significant predictor, $\beta=.765$, Wald=1.77, $p=.184$, this may have been due to insufficient power, because we were limited to analyzing only the relatively small number of subjects who advanced from Introductory Calculus for Engineers to Engineering Analysis I and who furthermore took all exams in Engineering Analysis I. Nonetheless, the odds ratio was 2.28,

Table 4 Number of subjects (%) receiving satisfactory and unsatisfactory grades in Engineering Analysis I

Group	Satisfactory (C or better)	Unsatisfactory (C− or worse)
Control	14 (48.3)	15 (51.7)
Experimental	17 (68.0)	8 (32.0)

indicating that students who had been in the experimental group were more than twice as likely as those who had been in the control group to earn a satisfactory grade. Given the practical significance of obtaining a satisfactory grade in a first engineering calculus course, we believe this result is worth noting, but it must be considered highly preliminary, given that the 95 % confidence interval ranged from .65 (a slightly negative effect of experimental group membership) to 8.1 (a strong positive effect).

Discussion

This study tested the efficacy of a pedagogical technique which cognitive psychologists have repeatedly recommended in recent years—namely, spacing of retrieval practice following initial acquisition of course material. We examined the real-world classroom performance of first-year engineering students who took a precalculus course and, in some cases, proceeded into a calculus course which depended heavily on precalculus knowledge. Our goal was to determine whether spaced retrieval practice can increase retention of a complex body of college-level mathematical knowledge in both the short- and long-term (i.e., within the precalculus course and into the calculus course, respectively). We obtained several encouraging findings.

Our test of relatively short-term retention was the cumulative final exam in Introductory Calculus for Engineers. At that time point, our study design allowed us to assess the effect of spaced retrieval practice both within- and between-subjects. Both comparisons revealed positive effects of spaced retrieval practice. Within the experimental group, subjects retained spaced objectives significantly better than massed ones. The mean difference in proportion correct between objective types was .03, which translates into a spacing-related improvement of one third of a letter grade. The effect of spacing appeared more potent when assessed between subjects. Subjects in the experimental group retained target objectives that were spaced significantly better than subjects in the control group retained those same target objectives when they were massed, and the difference in proportion correct between groups was quite practically significant (M_{diff}=.08), translating into more than half a letter grade.

Of importance, there are several reasons to attribute the f difference in retention of target objectives to manipulation of the IV rather than to some unidentified confound. First, group assignment was random and performed before the semester began by a member of the research team who had no contact with students, excluding the possibility of systematic bias. Second, groups were statistically indistinguishable on race, gender, high school GPA, and Math ACT—all variables that could plausibly be related to performance in a college mathematics course. Moreover, those variables were used as covariates in all analyses. Third, the groups did not differ significantly on the first unit exam in Introductory Calculus for Engineers, which was administered before subjects in the experimental group had much exposure to the spacing manipulation. This suggests that subjects in the two groups did not differ markedly in their baseline aptitude to learn precalculus. Together, these considerations favor the conclusion that it was manipulation of retrieval practice which produced superior retention of spaced objectives in the experimental group compared to massed objectives in the control group.

Our measures of longer-term retention were administered in Engineering Analysis I, the course following the one in which the temporal distribution of retrieval practice was manipulated. Between-groups comparisons again supported the efficacy of spaced retrieval practice. On the first unit exam in Engineering Analysis I, which covered material presented in

Introductory Calculus for Engineers, students who had been in the experimental group scored higher than students who had been in the control group. The difference in proportion correct was .06, equivalent to about half a letter grade. However, the difference only approached significance (p=.078). A larger, more robust effect was not necessarily to be expected, given that the unit exam was, naturally, proceeded by a review of material from Introductory Calculus for Engineers and this recent refresher might well have reduced differences between the groups in terms of what they retained from the previous semester. Nonetheless, the difference is entirely consistent with the one obtained on our longest-term measure of retention: the cumulative final exam in Engineering Analysis I. On that exam, students who had been in the experimental group a full semester previously scored significantly higher than students who had been in the control group. Once again, the difference in proportion correct was practically significant (M_{diff}=.10), translating into one full letter grade.

The groups we analyzed in Engineering Analysis I did not differ on any of the characteristics we controlled for. In addition, group scores did not differ on a unit exam in Engineering Analysis I that covered material wholly independent of any presented in Introductory Calculus for Engineers. Consequently, the between-group differences in Engineering Analysis I, like those in Introductory Calculus for Engineers, can be attributed with some confidence to the manipulation of retrieval practice in the precalculus course.

Our findings were anticipated by a literature demonstrating the mnemonic benefits of spaced retrieval practice in the laboratory and in a limited range of classroom settings, but they also extend that literature in crucial ways. Prior studies had not demonstrated an effect of spaced retrieval practice on mastery of complex bodies of mathematical knowledge such as precalculus and calculus. Nor had many studies been conducted in college classrooms where, a priori, it seemed conceivable that between-groups effects of spaced retrieval practice could be washed out by within-groups variability in students' prior knowledge, study habits, or any of a host of other individual-difference factors. And, finally, no studies had demonstrated benefits of spacing in the classroom that endured into subsequent semesters.

One other contribution of the present findings is that they support the value of spaced retrieval practice with a novel vulnerable population. Previously, spaced retrieval practice has been explored as a means of improving retention for individuals with memory impairment, ranging from impairment associated with normal healthy aging (e.g., Logan and Balota 2008) to more profound impairment associated with early stage dementia of the Alzheimer's type (e.g., Balota et al. 2006) and probable Alzheimer's disease itself (e.g., Cherry et al. 1999). Here, we were interested in a quite different sort of vulnerability. Recall our observation that Introductory Calculus for Engineers was designed for students who wished to pursue an engineering degree but who may have entered college with insufficient mathematical competency. Demonstrating the efficacy of spaced retrieval practice in a population that lacks preexisting strength in a particular domain is important for establishing the broad applicability of the technique. All other things being equal, mnemonic interventions that work for more diverse populations are preferable to those whose applicability is more limited (for discussion, see Dunlosky et al. 2013).

Although we believe that essentially all students who take Introductory Calculus for Engineers can be fairly characterized as at-risk for poor performance in the engineering program, it should be appreciated that our various analyses were conducted on only a subset of students who initially enrolled in the course (see Table 1). Students who not only completed one or both courses under investigation in this study but who furthermore took all exams may have been stronger and/or more motivated than those who did not. This possibility prevents us from concluding that spaced retrieval practice will necessarily benefit the very most vulnerable students in the population.

The present study was not designed to illuminate the theoretical mechanism by which spaced retrieval practice enhances knowledge retention beyond that afforded by massed practice. However, it is possible that study-phase retrieval is critical. When students attempted to answer quiz questions 1 or 3 weeks after first being introduced to a particular learning objective, they may have had to retrieve information they read about that objective in the course e-book. This could enhance memory for the textual information by serving as retrieval practice for it (e.g., Karpicke and Blunt 2011; Roediger and Karpicke 2006) and/or by providing an opportunity for the information to be bound to more diverse contextual cues (as noted by Cepeda et al. 2006). Moreover, the need to retrieve textual information might signal to students that they lack fluency with the material, prompting them to target the information for analytical processing (Alter et al. 2007). It is also conceivable that analytical processing would be directed at the corrected feedback which was provided when quiz questions were answering incorrectly, and this might also be expected to enhance retention (Pashler et al. 2005). Less study-phase retrieval, if any, might have been required to answer a second or third question about the same objective in quick succession on a single quiz.

In designing this study, we were greatly interested in the possibility that spacing retrieval practice of some objectives might increase performance on other, nonspaced objectives—an effect suggested by our within-subjects analysis, as just discussed. Our interest can be contrasted with Butler et al. (2014), who commented that topics in their engineering classroom study were "… designed to be relatively independent of each other in order to minimize the potential for learning from the intervention to affect learning from standard practice" (p. 334). To directly test for a "spill-over" effect of spaced retrieval practice, we utilized a design that allowed us to compare retention of massed objectives in an experimental group (with some spacing) to a control group (with no spacing). This comparison yielded tantalizing evidence of the anticipated effect. The experimental group performed better on massed objectives than did the control group and the difference fell just short of significance ($p = .058$). We are hesitant to make too much of a nonsignificant result, but tentatively venture the following possible explanation. The learning objectives we chose to target in Introductory Calculus for Engineers were highly interrelated and, by extending retrieval practice beyond the week in which objectives were initially acquired, we necessarily increased interleaving of spaced objectives (from previous weeks) with massed objectives (from the current week) in the experimental group. Interleaving has been shown to increase learning of interrelated mathematical operations (Rohrer 2009). Therefore, spaced retrieval practice may have increased mastery of massed objectives relative to mastery of the same objectives in the control group. We note that such an effect would be consistent with theoretical perspectives according to which interventions that increase retention of information also allow people to use that information more flexibly and perhaps master new concepts (Carpenter 2012; Karpicke 2012). If future research can establish that spaced retrieval practice has a genuine "side effect" of increasing mastery of nonspaced information, it would further strengthen the case for incorporating the technique into educational practice.

It bears noting that we do not know how aware students were of the manipulation in this study and we do not know whether it affected students' study habits. Regarding awareness, there are two questions. One, did students in the experimental group notice that some quiz questions they received were massed and others were spaced? Two, did students in either group realize that some of their classmates received slightly different quizzes? We have no evidence—formal or informal—that either type of awareness occurred. There were 214 learning objectives in Introductory Calculus for Engineers and quiz questions targeting all of them were massed, save for 24 (11.2 %) in the experimental group. While some students may have noticed the infrequent exceptions to the rule, it is feasible that others did not. Quizzes were administered digitally at the time and place of students' choosing, meaning that many students probably took them by themselves, reducing the likelihood that members of different groups would chance to communicate afterwards and discover dissimilarity in quiz content. Ultimately, the safest assumption is probably that some students became aware of the manipulation and some did not, with awareness being more likely among students exposed to the manipulation within-subjects (i.e., in the experimental group).

Students in the experimental group who became aware of the spacing manipulation may have adopted different study habits than students in the control group. Specifically, students who anticipated having to solve problems that covered particular objectives a week or more after initial acquisition may have studied those objectives more and/or extended their study beyond the week of initial acquisition (i.e., spontaneously spacing their study). These behaviors, if they occurred, could have contributed to the experimental group's superior retention of course material. While laboratory studies can eliminate the possibility of differential study between spaced and massed conditions by tightly controlling subjects' access to to-be-learned materials, classroom studies cannot do so, at least not without sacrificing a great deal of ecological validity. Because the precise mechanism or combination of mechanisms by which spaced retrieval practice improves classroom performance is of interest, future research could attempt to measure students' amount and distribution of study time.

Conclusions

Our findings strongly support cognitive psychologists' optimistic predictions that spaced retrieval practice can have a genuinely meaningful impact on college education. Demonstrating the efficacy of spaced retrieval practice in precalculus and calculus classes for engineers is especially encouraging because it suggests the technique may also help prepare students for careers in other math-intensive STEM fields, such as chemistry, physics, and statistics. We point out that spaced retrieval practice is a no-cost intervention that does not involve adding anything to courses or requiring students to do anything they would not ordinarily do. It merely involves modifying the temporal distribution of existing activities. This may make spaced retrieval practice particularly attractive to those educators concerned about increasing workload for themselves or their students, as can happen with certain other possible interventions, such as increasing the number of retrieval exercises (e.g., Lyle and Crawford 2011) or requiring students to view feedback (e.g., Butler et al. 2014). While we would not discourage educators from considering every empirically validated intervention, including combinations of techniques, it is heartening to know that even a single, relatively minimal change to the status quo has the potential to dramatically increase student success.

Acknowledgments This research was supported by a National Science Foundation Improving Undergraduate STEM Education Award.

Compliance with Ethical Standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964

Helsinki declaration and its later amendments or comparable ethical standards. The Institutional Review Board at our university waived the requirement for subjects in this research to provide informed consent. However, as stated in the manuscript, students were given the option to decline to have their performance data included in our analyses. No student declined. This article does not contain any studies with animals performed by any of the authors.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*, 569–576.

Bacon, D. R., & Stewart, K. A. (2006). How fast do students forget what they learn in consumer behavior? A longitudinal study. *Journal of Marketing Education, 28*, 181–192.

Balota, D. A., Duchek, J. M., Sergent-Marshall, S. D., & Roediger, H. L., III. (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging, 21*, 19–31.

Beanland, D. G. (2010). *Challenges and opportunities facing the education of engineers*. Retrieved from https://www.engineersaustralia.org.au/sites/default/files/shado/Divisions/Victoria%20Division/Groups/Senior%20Engineers%20Group/seg_march2010.pdf.

Butler, A. C., Marsh, E. J., Slavinsky, J. P., & Baraniuk, R. G. (2014). Integrating cognitive science and technology improves learning in a STEM classroom. *Educational Psychology Review, 26*, 331–340.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*, 279–283.

Carpenter, S. K. (2014). Spacing and interleaving of study and practice. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying the science of learning in education: infusing psychological science into the curriculum* (pp. 131–141). American Psychological Association.

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.

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, 19*, 1095–1102.

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

Cherry, K. E., Simmons, S. S., & Camp, C. J. (1999). Spaced retrieval enhances memory in older adults with probable Alzheimer's disease. *Journal of Clinical Geropsychology, 5*, 159–175.

Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: twelve years of cognitive psychology. *Journal of Experimental Psychology. General, 120*, 395–409.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology, 21*, 919–940.

Cull, W. L. (2005). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235.

Delaney, H. D., & Maxwell, S. E. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behavioral Research, 16*, 105–123.

Delaney, P. F., Verkoeijen, P. P. J. L., & 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.

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*, 4–58.

Ebbinghaus, H. (1964). *Memory: a contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.). New York: Dover Publications. (Original work published 1885).

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*(40).

Goosens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28*, 135–142.

Hieb, J. L., Lyle, K. B., Ralston, P. A. S., & Chariker, J. (2015). Predicting performance in a first engineering calculus course: implications for interventions. *International Journal of Mathematical Education in Science and Technology, 46*, 40–55.

Kamuche, F. U., & Ledman, R. E. (2005). Relationship of time and learning retention. *Journal of College Teaching & Learning, 2*, 25–28.
Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology. General, 138*, 469–486.
Karpicke, J. D. (2012). Retrieval-based learning: active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*, 157–163.
Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772–775.
Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of experimental psychology: learning, memory, and cognition, 33*, 704–719.
Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.
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.
Landauer, T. K., & Eldridge, L. (1967). Effect of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology, 75*, 290–298.
Logan, J. M., & Balota, D. A. (2008). Expanded versus equal interval spaced retrieval practice: exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, & Cognition, 15*, 257–280.
Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology, 38*, 94–97.
Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis, 35*, 105–123.
McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*, 360–372.
McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied, 20*, 3–21.
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.
Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (*NCER 2007–2004*). Washington: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. **Retrieved from http://ncer.ed.gov**.
Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science, 26*, 784–793.
Pearson, W., & Miller, J. D. (2012). Pathways to an engineering career. *Peabody Journal of Education: Issues of Leadership, Policy, and Organizations, 87*, 46–61.
Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review, 25*, 523–548.
Rickard, T. C., Lau, J. S.-H., & Pashler, H. (2008). Spacing and the transition from calculation to retrieval. *Psychonomic Bulletin & Review, 15*, 656–661.
Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.
Roediger, H. L., III, & 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*, 242–248.
Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40*, 4–17.
Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin, 140*, 1432–1463.
Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review, 21*, 1516–1523.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366.
Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*, 763–767.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*, 837–848.

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.

Thomas, M. S. C., Annaz, D., Ansari, D., Serif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research, 52*, 336–358.

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

White House (n.d.). *Educate to innovate*. Retrieved from https://www.whitehouse.gov/issues/education/k-12/educate-innovate.