Retrieval Practice and Spacing in an Engineering Mathematics Classroom: Do the Effects Add Up?

Campbell R. Bego, Keith B. Lyle
Department of Psychological and Brain Sciences
University of Louisville
Louisville, KY, USA
carigh01@louisville.edu, kblyle01@louisville.edu

Patricia A. Ralston, Jeffrey L. Hieb
Department of Engineering Fundamentals
UofL, J.B. Speed School of Engineering
Louisville, KY, USA
parals01@louisville.edu, jlhieb01@louisville.edu

***Abstract*—Mathematics learning is critical in STEM degrees. In engineering specifically, advanced courses depend on the derivation and application of higher-order equations, and long-term retention of early mathematical concepts is mandatory. Cognitive science research has shown that learning and memory can be improved with simple manipulations of material retrieval. One major finding has been that increasing retrieval practice improves material retention more than restudying the material (the retrieval-practice effect). Independently, spacing retrieval over time is known to improve material retention (the spacing effect). To date, no research has studied these two effects in conjunction. This NSF-funded study investigated retrieval practice and spacing in an engineering mathematics classroom. Separately, increasing retrieval practice and increasing spacing both improved final exam performance. In combination, they produced the greatest learning gains.**

Keywords— Memory, Spacing, Testing, Retrieval Practice, Engineering Mathematics, Educational Research

I. Introduction

This research paper is part of a study funded by the National Science Foundation (NSF) at the University of Louisville J.B. Speed School of Engineering. The ongoing study is investigating ways to improve engineering mathematics instruction in the classroom. Calculus poses a challenge for many STEM students, with many classes suffering from pass rates below 50%. In fact, one prominent Australian engineering educator claims "the biggest factor contributing to the failure of engineering students is inadequate competence in mathematics" [1]. Researchers Pearson and Miller [2] 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 college calculus courses taken, suggesting that mathematics performance is critical for student retention in engineering. Methodologies that lead to improved performance in mathematics classes therefore have the potential to positively impact student learning in later courses, and such interventions may eventually lead to improvements in the historically low student retention rates in engineering.

Research in cognitive psychology suggests that learning and memory could be enhanced by changing how often and on what temporal schedule students are required to retrieve and apply mathematics knowledge. For example, increasing the number of times information is retrieved has been shown to improve retention [3]. In fact, retrieving a piece of information from memory bolsters the long-term retention of that information more than does restudying the same piece of information (e.g. [4], [5]). The superiority of retrieval over restudy is known as the retrieval practice effect (or the testing effect). This effect of retrieval practice on memory have been studied extensively, and the results are highly reliable in many domains [6]. Classroom studies have shown that, when instructors implement methods that require students to practice retrieval more than they would on their own, students retain more course content and earn higher grades (e.g., [7]–[9])

Additionally, retention of material improves when multiple retrievals are spaced out over time versus being temporally massed ([10], [11]). This is referred to as the spacing effect. Like the retrieval practice effect, the spacing effect has been extensively studied and results show consistent and robust effect sizes. Research suggests that college students are not cognizant of the value of spacing ([11], [12]) and therefore they may not utilize it spontaneously in self-initiated study practices. Nor do common educational practices promote much spacing. In many classes, material is learned and tested in discrete units. Students are required to retrieve unit content—on homework, quizzes, tests, or all of these—but retrieval opportunities rarely extend beyond the end of the unit, when the focus shifts to learning and retrieving new information. In other words, retrieval practice is commonly "massed" in a short temporal window. A form of spacing, called interleaving, has been shown to increase retention of mathematics knowledge in the laboratory [13], [14].

Recently with support from NSF DUE-IUSE award 1431544, and to our knowledge for the first time, it was shown that spaced retrieval practice in a mathematics course for engineers increased students' retention of course content [15]. Of importance, the positive effect of spacing on retention was also evident in a subsequent mathematics course. This next course built heavily on concepts to which the spacing manipulation was applied. Students whose retrieval practice had been spaced in the first course performed better on the final exam in the subsequent course than students whose practice had been massed, suggesting greater retention among the former than the latter.

Cognitive psychologists have enthusiastically recommended that educators incorporate more retrieval practice and more spacing into their pedagogical practices ([3], [16]–[18]). Implicit in these recommendations is the idea that, since both increasing the amount and increasing the spacing of retrieval practice are known to benefit human memory, educators should do both, if possible. For example, engineering mathematics educators could both increase the number of quiz questions covering a particular learning objective and increase the questions' spacing.

Making both changes might sound reasonable, but that recommendation critically assumes that increasing the amount of retrieval practice is additive with the effect of increasing the spacing. In simplest terms, the assumption here is that, because increasing the amount of retrieval practice is good and increasing the spacing of practice is good, increasing both in combination is better than increasing either one alone. However, to our knowledge, such an additive effect has never been empirically established. This is an important limit to our basic knowledge about the retrieval-practice and spacing effects and it also limits what can be justifiably recommended to educators. Another important limitation is that we do not know how the magnitude of the retrieval-practice effect directly compares to the magnitude of the spacing effect.

This study investigates whether the retrieval-practice and spacing effects are additive in the engineering mathematics classroom, and which, if either, is superior. We manipulated the number and temporal distribution of quiz questions in a preparatory math course for first-year engineering students. This course is called *Introductory Calculus for Engineers,* and it is taken by students who would otherwise likely struggle in a first-year calculus-based engineering mathematics course. We assessed the independent and additive effects of our interventions on course performance, specifically performance on the final exam. Evidence that retrieval practice and spacing are additive would support additional investigation into additive effects of other research-based instructional strategies, and may encourage faculty to more aggressively embrace and incorporate multiple well-researched instructional strategies.

II. Methodology

This study compared spacing and increased retrieval practice effects in a 2x2 within-subjects experimental design. The manipulation was implemented in *Introductory Calculus for Engineers* at the University of Louisville's J.B. Speed School of Engineering in Fall 2016, and utilized MyMathLab®, an online media and homework system developed by textbook publisher Pearson. The study was approved by our Institutional Review Board (16.0492).

A. Participants

All students who registered for *Introductory Calculus for Engineers* in Fall 2016 were participants in the study ($N = 139$). Data were only included in the analysis, however, if the student took all quizzes as well as the final exam ($N = 62$).

This research is sponsored by the National Science Foundation: NSF IUSE EHR 1609290.

B. Course Format and Materials

The *Introductory Calculus for Engineers* course was designed to give students a solid background in algebra, trigonometry, analytic, geometry, and an introduction to differential calculus. 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, tested themselves as part of structured sequences called study plans, and took two quizzes (one per week) and a unit exam. Students were required to complete study plans prior to quizzes, and quizzes were required prior to taking exams. However, the students were allowed one exemption to take an exam without having taken a quiz. Most students used this exemption at the beginning of the semester while working out the self-paced schedule. There were no lectures, however students attended weekly class meetings (150 minutes per week) in which the course instructor and teaching assistants answered questions and administered collaborative problem-based activities. The instructor also reviewed the schedule of upcoming assignments, discussed administrative issues, and discussed the relevance of course topics to engineering.

Amount of retrieval practice was manipulated by varying the number of quiz questions targeting a given learning objective. For the purpose of the study, we focused on 32 specific target objectives (a subset of the total in the class). Four of these objectives were introduced each week over the course of eight weeks. An example learning objective is *Simplifying Exponential Expressions Involving Rational Exponents*. Half the objectives were covered by three quiz questions (less practice) and half by six (more practice). Spacing was manipulated by varying the placement of questions on quizzes. For half the objectives, all questions (whether three or six) appeared on a single quiz (massed practice). For the remaining objectives, one-third of the questions (either one or two) appeared on each of three different quizzes, administered in different weeks (spaced practice). The spacing schedule used in this study mirrors the spacing in our previous study [15]: one question was given the week of instruction, one question a week later, and the last question two weeks after that. This resulted in four conditions: *less practice, massed* (3 questions in one week), *less practice, spaced* (3 questions over four weeks), *more practice, massed* (6 questions in one week), and *more practice, spaced* (6 questions over four weeks).

Each learning objective introduced in a given week was assigned to a different one of the four conditions, allowing for a within-subjects comparison of performance across the four conditions. Assignment of learning objective to condition was counterbalanced, meaning that each objective was quizzed in each condition for an equal number of students. The experiment ran from week 1 to week 11 of the semester, introducing new target objectives in quizzes 1-8 and finishing the spaced questions in week 11.

All quizzes were created and distributed using Pearson's MyMathLab®. Six questions were prepared for each target objective using the available online materials. For the *less practice, massed* condition, the first three questions were given in the first week. The same questions were used in the *less* *practice, spaced* condition over four weeks. For the *more practice, massed* conditions, all six questions were assigned in the first week, and the same questions were used in the *more practice, spaced* condition over four weeks. For each week of the semester, four counterbalanced quizzes were constructed. Each quiz had the target objectives spaced according to the counterbalanced conditions of the objectives by group. Because of the counterbalanced design, the four quizzes had the same number of questions. In total, 45 quizzes were created for this study (11*4 counterbalanced + 1 quiz for week 12). Facility with a tool such as MyMathLab® (or similar online system capable of grading, capturing, and storing the necessary data) is necessary for implementation of this project's research design. Questions were chosen and reviewed by multiple researchers and the final quizzes were checked by the PI.

C. Procedures

The 32 target objectives were selected at the start of the project, approximately three months prior to implementation. These target objectives were selected from a much longer list of course objectives and were selected to be distributed evenly over the first four units, or eight weeks, of the class. Four target objectives were chosen each week. For each target objective, six distinct questions of moderate difficulty were identified to test each objective. The first three questions were administered in all conditions. Questions 4-6 were administered only in the *more-practice, massed* and *more-practice, spaced* conditions. Four different MyMathLab® "courses" were then created with twelve quizzes in each course. Across courses, assignment of target objectives was counterbalanced. Quizzes were generated during the summer prior to the Fall 2016 semester. Quizzes were then administered once a week during a 48-h window that was the same for all students, regardless of group assignment.

During the semester, quiz data were downloaded from MyMathLab®. Personal information was stripped from the data, and a non-identifiable Research ID (RID) was used to organize the data. Data was combined in a MySQL database and all quiz data were preserved for future analysis. Additionally, performance on the Final Exam was downloaded from MyMathLab®. Only the questions that were from target objectives were preserved in the dataset.

The final exam was administered at the end of the semester, also using MyMathLab®. The final exam included questions that covered all target objectives, as well as non-target objectives, but only target objectives are of interest to this study. The number of questions per condition was balanced in the final exam (1 or 2 questions per objective resulting in ~11 questions per the 8 objectives in each condition). For example, student group 1 saw objectives 1, 5, 9, 13, 17, 21, 25, and 29 in the *less-practice, massed* condition, and there were 11 questions covering those objectives on the final exam. Student performance was averaged across questions for each condition.

D. Analyses

A 2 (*less practice, more practice*) x 2 (*massed, spaced*) repeated-measures analysis of variance (ANOVA) was performed. If increasing retrieval practice from three to six questions enhances retention of mathematical knowledge, then there should be a significant main effect of amount of practice. Specifically, proportion correct should be greater in the six-question conditions than in the three-question conditions. The mean difference between 6- and 3-question conditions would be a practically meaningful estimate of the impact of increasing the amount of retrieval practice. We also examined partial eta squared for effect size.

If increasing spacing of retrieval practice enhances retention of mathematical knowledge, then there should be a significant main effect of temporal distribution. Specifically, proportion correct should be greater in spaced conditions than massed ones. The mean difference between spaced and massed conditions would be a practically meaningful estimate of the impact of increasing spacing. We compared the mean difference between the spaced and massed conditions to the mean difference between the 6- and 3-question conditions as well as the two effect sizes to determine which single intervention (increasing spacing or increasing the amount of retrieval practice) had a larger impact on retention of mathematical knowledge.

An additive effect of increasing the amount and spacing of retrieval practice would be revealed by the existence of two significant main effects (of the form just described) without a statistically significant interaction. If this pattern were to emerge, proportion correct would be greatest in the six-spaced condition and it would be interesting to directly compare it to proportion correct in the three-massed condition, which in some respects represents the status quo in common educational practice (i.e., relatively little retrieval practice and little spacing).

III. Results

Results from the 2 (*less practice* versus *more practice*) x 2 (*massed* versus *spaced*) within-subjects repeated-measures ANOVA are reported here.

As mentioned above, 62 participants took all quizzes and the final exam. The distribution of participants in the counterbalancing groups was as follows: 18, 14, 14 and 16. Analyses were performed using the unequal group sizes and then performed again with a smaller subset of participants to equalize the counterbalanced groups ($N = 14$ of each for a total of $N = 56$). Four participants were removed from one sub-group and two from another. Removed participants were those with the largest random RID numbers. Results of both analyses are reported here.

A. ANOVA - all participants (N = 62)

The main effect of spacing was significant, $F(1, 61) = 10.536$, $p = 0.002$, $\eta^2 = 0.147$, while the main effect of retrieval practice fell just short of significance, $F(1, 61) = 3.894$, $p = 0.053$, $\eta^2 = 0.060$. There was no interaction between retrieval practice and spacing. Means are reported in Table 1 and shown in Figure 1.

B. ANOVA - subset of participants (N = 56)

Results were very similar in a subset of participants (equal N in each counterbalanced group). The main effect of spacing was significant, $F(1, 55) = 11.838$, $p = 0.001$, $\eta^2 = 0.177$, and the main effect of retrieval practice reached conventional significance, $F(1, 55) = 4.891$, $p = 0.031$, $\eta^2 = 0.082$. There was no interaction between retrieval practice and spacing. Means are reported in Table 2 and shown in Figure 2.

Table 1:
Final Exam Performance, All Participants

Condition	Mean	Std. Dev.	Std. Error	N
Less practice, massed	0.624	0.16133	0.00260	62
Less practice, spaced	0.666	0.18777	0.00303	62
More practice, massed	0.655	0.16866	0.00272	62
More practice, spaced	0.701	0.18074	0.00292	62

Table 2:
Final Exam Performance, Subsample of Participants

Condition	Mean	Std. Dev.	Std. Error	N
Less practice, massed	0.631	0.16367	0.00292	56
Less practice, spaced	0.667	0.19272	0.00344	56
More practice, massed	0.657	0.17490	0.00312	56
More practice, spaced	0.720	0.17623	0.00315	56

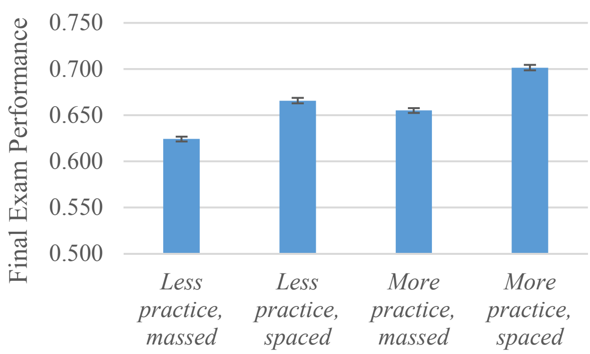

Figure 1: Mean final exam performance by condition for all participants ($N = 62$).

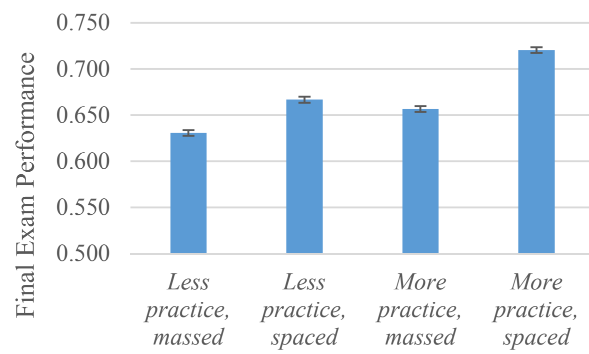

Figure 2: Mean final exam performance by condition for a subset of participants ($N = 56$).

C. *Comparison of Means*

Examining results from all participants and collapsing across levels of retrieval practice, the mean of the *massed* conditions was 0.640 and the mean of the *spaced* conditions was 0.684, indicating that spacing increased performance by 4.4%. Collapsing across levels of the spacing factor, the mean of the *less practice* conditions was 0.645 and the mean of the *more practice* conditions was 0.678, revealing a 3.3% performance increase due to more retrieval practice. Corresponding numbers for the subset analysis were 5.0% and 4.0%, respectively.

IV. Discussion

The results demonstrate a robust, significant main effect of spacing, regardless of whether the sample is full or cut. Students performed statistically better on the final exam on objectives that were presented to them with spacing as opposed to massed. The effect size was $\eta^2 = 0.147$ for all participants, and for the subset, $\eta^2 = 0.177$, which is classified by Cohen [19] as small but not trivial. Mean final exam performance in the *spaced* conditions was 4.4% and 5.0% higher than in the *massed* conditions (all participants and subset respectively), which is equivalent to at least half of a letter grade. This impact on learning with no additional effort from students or teachers, other than planning, is a tangible benefit. These results replicate those of the previous NSF study, demonstrating that spacing reliably improves retention of mathematics knowledge over the course of a semester.

Compared to the effect of increased spacing, the effect of increased retrieval practice was less clear. In the full sample, the effect of increased retrieval practice fell just short of significance ($p = .053$), although the effect was conventionally significant in the analysis conducted on a slightly reduced sample. Regardless of analysis, the effect size (Full: $\eta^2 = 0.060$, Subset: $\eta^2 = 0.082$) was notably lower than the effect size of spacing. The average mean in final exam performance was higher in the *more practice* than the *less practice* groups (Full: +3.3%, Subset: +4.0%) which is similar but less than the difference due to spacing. The smaller effect size and smaller mean difference as well as the higher p-values indicate that the effect of increased retrieval practice is smaller and less robust than the spacing effect.

In terms of classroom application, instructors may be less likely to increase spacing than to increase retrieval practice, because increasing spacing arguably requires more careful planning than does adding more questions. Increasing spacing requires instructors to develop a spacing schedule and to remember to adhere to it days and weeks after material is initially tested. However, technology such as Pearson's MyMathLab® and other tools may make implementation of spacing easier. With respect to other interventions like private tutoring, reducing class sizes, and study strategies training, adding spaced questions throughout a curriculum is an extremely low-cost alternative.

Importantly, the two main effects indicate that the effects are additive. The largest effect was seen in the condition in which more questions were administered and spacing was implemented. The average difference in final exam performance between the *more practice, spaced* condition and the baseline *less practice, massed* condition was 8.3%. This constitutes an entire letter grade. Combining these two educational practices therefore produced a marked increase in student learning.

These results are even more impressive given that we could in no way control students' self-initiated study habits in an actual classroom. For example, many students presumably studied intensively for the final exam, treating many target objectives the same, despite differential treatment as part of the experiment. Plausibly, this could have washed out effects of our experimental manipulations, but this appears not to have happened. This suggests the benefit of spacing (and to a lesser extent the benefit of retrieval practice) is robust in the face of students' real-world study behaviors.

The math course *Introductory Calculus for Engineers* is geared toward students with lower mathematics preparation and performance levels. Improving students' performance by 8% is a large gain in the right direction.

V. Conclusions

This paper discusses the results of an in-class manipulation of retrieval practice and spacing. Results showed that the two effects are additive, and that the spacing effect is more robust. The results of this research will permit more refined recommendations regarding the use of retrieval practice and spacing in foundational mathematics courses across the STEM curriculum. The results would substantially increase the evidentiary base from which to stipulate best practices in mathematics education. This could reduce the number of students transferring out of STEM majors due to insufficient mathematical knowledge.

This research also shows that further investigation is warranted on the implementation of psychological effects in the classroom, and that it is a fruitful avenue of research moving forward.

References

[1] D. G. Beanland, "Challenges and Opportunities Facing the Education of Engineers-Address to Victoria Division of Engineers Australia Seg Meeting." Melbourne, 2010.

[2] W. Pearson Jr and J. D. Miller, "Pathways to an engineering career," *Peabody J. Educ.*, vol. 87, no. 1, pp. 46–61, 2012.

[3] H. L. Roediger and M. A. Pyc, "Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice," *J. Appl. Res. Mem. Cogn.*, vol. 1, no. 4, pp. 242–248, 2012.

[4] J. D. Karpicke and H. L. Roediger, "The critical importance of retrieval for learning," *Science,* vol. 319, no. 5865, pp. 966–968, 2008.

[5] H. L. Roediger III and J. D. Karpicke, "The power of testing memory: Basic research and implications for educational practice," *Perspect. Psychol. Sci.*, vol. 1, no. 3, pp. 181–210, 2006.

[6] C. A. Rowland, "The effect of testing versus restudy on retention: A meta-analytic review of the testing effect." *Psychol. Bull.* vol. 140, no. 6, 2014.

[7] F. C. Leeming, "The exam-a-day procedure improves performance in psychology classes," *Teach. Psychol.*, vol. 29, no. 3, pp. 210–212, 2002.

[8] K. B. Lyle and N. A. Crawford, "Retrieving essential material at the end of lectures improves performance on statistics exams," *Teach. Psychol.*, vol. 38, no. 2, pp. 94–97, 2011.

[9] M. A. McDaniel, J. L. Anderson, M. H. Derbish, and N. Morrisette, "Testing the testing effect in the classroom," *Eur. J. Cogn. Psychol.*, vol. 19, no. 4–5, pp. 494–513, 2007.

[10] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis.," *Psychol. Bull.*, vol. 132, no. 3, p. 354, 2006.

[11] N. Kornell, "Optimising learning using flashcards: Spacing is more effective than cramming," *Appl. Cogn. Psychol.*, vol. 23, no. 9, pp. 1297–1317, 2009.

[12] J. M. Logan, A. D. Castel, S. Haber, and E. J. Viehman, "Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal," *Metacognition Learn.*, vol. 7, no. 3, pp. 175–195, 2012.

[13] D. Rohrer and K. Taylor, "The effects of overlearning and distributed practise on the retention of mathematics knowledge," *Appl. Cogn. Psychol.*, vol. 20, no. 9, pp. 1209–1224, 2006.

[14] D. Rohrer and K. Taylor, "The shuffling of mathematics problems improves learning," *Instr. Sci.*, vol. 35, no. 6, pp. 481–498, 2007.

[15] R. F. Hopkins, K. B. Lyle, J. L. Hieb, and P. A. S. Ralston, "Spaced Retrieval Practice Increases College Students' Short-and Long-Term Retention of Mathematics Knowledge," *Educ. Psychol. Rev.*, vol. 28, no. 4, pp. 853–873, 2016.

[16] H. Pashler, P. M. Bain, B. A. Bottge, A. Graesser, K. Koedinger, M. McDaniel, and J. Metcalfe, "Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004.," *Natl. Cent. Educ. Res.*, 2007.

[17] S. K. Carpenter, "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, 2014.

[18] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham, "Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology," *Psychol. Sci. Public Interes.*, vol. 14, no. 1, pp. 4–58, 2013.

[19] J. Cohen, "Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates," *Hillsdale, NJ*, pp. 20–26, 1988.