Using SQL to query the difficulty imposed by spaced retrieval in engineering mathematics

Jeremy R. Boyd
Department of Engineering Fundamentals
J. B. Speed School of Engineering
University of Louisville
jeremy.boyd@louisville.edu

Campbell R. Bego*
Department of Engineering Fundamentals
J. B. Speed School of Engineering
University of Louisville
campbell.bego@louisville.edu

Osvaldo Garcia
Department of Engineering Fundamentals
J. B. Speed School of Engineering
University of Louisville
osvaldo.garcia@louisville.edu

Patricia A. S. Ralston
Department of Engineering Fundamentals
J. B. Speed School of Engineering
University of Louisville
patricia.ralston@louisville.edu

Jason C. Immekus
Department of Educational Leadership, Evaluation and Organizational Development
College of Education
University of Louisville
jason.immekus@louisville.edu

Keith B. Lyle
Department of Psychological and Brain Sciences
College of Arts & Sciences
University of Louisville
keith.lyle@louisville.edu

***Abstract*—This WIP Research Paper investigates the temporal nature of the difficulty imposed by spacing. Spaced retrieval practice is an evidence-based strategy for improving memory and consists of asking multiple questions on a topic with intermittent delays. Spacing is often thought to impose difficulty by making questions harder to answer. However, this difficulty may be desirable, since spacing ultimately improves memory. In this paper, we (a) outline an implementation of spaced retrieval practice in an engineering mathematics classroom, (b) describe the development of an SQL database to organize and manage the large and complex dataset, and (c) discuss a brief but interesting dive into the rich data we have collected. Statistical analyses revealed that, when three questions targeting the same topic are spaced over multiple quizzes, versus being massed on a single quiz, students are less likely to answer the first and second questions correctly. Spacing does not affect students' ability to answer the third questions. This suggests that spacing may impose difficulty when students are first learning to perform mathematical operations, rather than when they are trying to retrieve memories of how to perform those operations.**

***Keywords*—spaced retrieval practice, spacing, desirable difficulty, engineering mathematics, SQL**

I. Introduction

Performance in mathematics is critical for student persistence and graduation in engineering [1]–[4]. It is therefore desirable to apply evidence-based instructional practices in mathematics classrooms and assess whether they have the intended impacts. One technique that has recently been shown to be effective in improving engineering mathematics performance is *spaced retrieval practice* [5], [6]. To implement spaced retrieval practice, instructors assign multiple recall exercises on the same topic over time. This method capitalizes on the benefits of the *testing effect* (i.e., answering questions is better for learning than restudying; [7]) and the *spacing effect* (i.e., spacing out content temporally is better for learning than presenting it all at one time; [8], [9]). Spaced retrieval practice has been shown to provide robust improvement in long-term memory.

Recent studies have shown that, although spacing increases performance on final outcome measures, it also reduces performance on initial practice questions [6], [10]. These short-term costs and long-term benefits constitute an example of *desirable difficulty*, which is the idea that challenges during initial learning can sometimes improve long-term retention [10], [11]. Spacing retrieval, versus retrieving repeatedly in a short span of time, can pose a challenge because it is often difficult to retrieve information that has not recently been accessed. This difficulty may be desirable, however, because it requires activation of retrieval processes that are required for retrieval in the long-term.

Currently in the second year of a 3-year project funded by the National Science Foundation, we collected student performance data from an implementation of spaced retrieval practice in an engineering mathematics course (Calculus I with engineering applications). In this study, we manipulated retrieval practice for twenty-four target learning objectives and selected questions that assess these learning objectives. Three algorithmic variants of each question were administered in quizzes over the course of one semester. In the *spaced* condition, one variant was asked in each of three quizzes. In the *massed* condition, all three variants were asked on the same quiz.

In a recent ASEE paper, we presented a preliminary analysis of the difficulty imposed by spacing [12]. We compared conditions by averaging student performance across the three question instances in each condition. Students were slightly, but significantly, less accurate when question instances were spaced versus massed—a difference of 2.61%. However, using the average may have underestimated the magnitude of spacing-induced difficulty. In the spaced condition, the first question was asked on the same quiz as the massed questions, i.e., with no delay. Looking at the first, second, and third question instances

This project was funded by the National Science Foundation, Award # 1912253.

separately might reveal larger performance differences between spaced and massed questions.

In the current study, we considered each question instance separately. Our research question was:

What is the temporal nature of the difficulty imposed by spacing? I.e., What was student performance on the first, second, and third question instance in the spaced condition, and how do those values compare to performance in the massed condition?

The primary hypothesis was that performance would decrease over time in the spaced condition due to the added difficulty of retrieval.

In addition, the second purpose of this paper was to describe an SQL database for collecting and organizing study data, which enabled us to run multiple analyses quickly and easily. The methodology section includes a detailed description of the database and how it aided data analysis.

II. Methodology

A. Participants

Participants ($N = 183$) were students who (a) were enrolled in Engineering Analysis I (Calculus I with engineering applications) at the JB Speed School of Engineering at the University of Louisville in Fall 2020, (b) completed all five practice quizzes, and (c) had no computer or internet-access errors that would have interfered with our experimental manipulations.

B. Procedures

First, the lead instructor of Engineering Analysis I selected 24 target learning objectives and corresponding quiz questions from an online learning platform. The target learning objectives and corresponding questions were selected to fit within the study schedule: eight learning objectives introduced in the first three weeks of the semester, eight in weeks four to five, and eight in weeks six to seven.

We then built two sets of five practice quizzes using the target learning objectives and associated questions. For the 5 practice quizzes, questions were assigned to either a massed or spaced condition. In the massed condition, a question was assigned three times on the same quiz. The online learning platform generated algorithmic variants of the questions such that students saw three similar questions that differed only in coefficients, variable names, or other numerical values. In the spaced condition, questions were asked on three consecutive quizzes, also with random algorithmic variants. In each quiz, questions were presented in random order.

The research design was within-subjects: half of the objectives were assigned to the massed condition, and half were assigned to the spaced condition. Thus, each student answered questions in both massed and spaced conditions. Assignment of objective to condition was counterbalanced, meaning each objective was assigned to the spaced condition for half the students and to the massed condition for the other half. To keep track of this counterbalancing, the two halves of the class were arbitrarily labeled Group A and Group B.

Quizzes were administered on weekends from Friday 1:00 pm to Sunday 11:59 pm, following weeks three, five, seven, nine, and eleven. A final quiz was administered on the last day of class (in week 14) as a criterial test of student learning. This paper considers only the first five practice quizzes and the difficulty imposed by spacing.

C. Materials

Study materials are not described in this WIP paper. Please see the recent ASEE publication for these details [12], or ask the corresponding author for additional information.

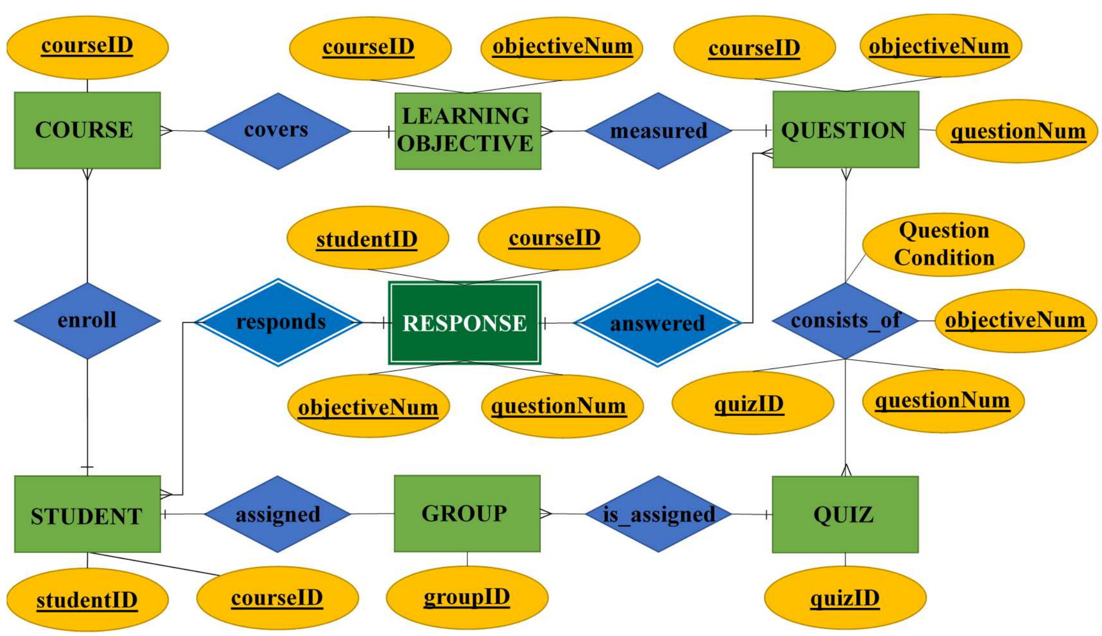

Fig. 1: A simplified Entity-Relationship Diagram (ERD) of the database used to store and retrieve data for the current spaced retrieval practice study. Beginning at the lower left hand corner and travelling counter-clockwise, the ERD illustrates that **Students** were *assigned* to a **Group**, each of which was *assigned* a set of **Quizzes**. **Quizzes** *consisted of* **Questions** that were asked in either a massed or spaced condition. The center of the figure illustrates' storage of **Students' Responses** to the **Questions**. The upper-left corner of the figure was for the multiple courses involved in this grant.

D. Storing and organizing the data

The full 3-year project included similar procedures in ten courses, with a total of 1315 participants. Because data organization is complex and of great importance in a study of this size, an SQL database was developed using MySQL software.

A postdoctoral researcher and two undergraduate research assistants first created an entity-relationship diagram to organize related fields (Fig. 1). In this type of diagram, each rectangle is an "entity" and each diamond is a "relationship," both of which store information in tables. The `STUDENT` table, for example, contained a de-identified student ID, some basic demographic information, and the course the student was taking. The `GROUP` entity table contained the two different group labels, and the `ASSIGNED` table held student ID and the group assignment. The database structure allowed us to populate and search the data in systematic ways.

We also created Views of the data that combine the tables to provide data access on a macro level. For example, a view called '`FiveQuizzes NoErrors`' included only data for students who were to be included in this study. The view contained information about the students, individual questions, and student performance on the questions. Specifically, it included *studentID*, *race*, *gender*, *course*, and *group* variables from the `STUDENT` and `ASSIGNED` tables, paired with *quiz number*, *objective number*, *question number*, and *score* from the `RESPONSE` table. Also in this view, we pulled in information from the `CONSIST_OF` table to get the *question condition* for each student for each question, which indicated whether a question was spaced or massed. This is where the power of database organization helped the most. Because of our within-subjects design with counterbalancing, students in the different groups had different questions in the two conditions. The view we created of student performance had *question condition* as an automatically populated field. In a typical Microsoft Excel-based analysis process, these steps would have to be done manually, for each quiz for each group of students. This process not only takes a great amount of time, but also increases the risk of reduced data quality due to human error.

E. Data Analysis

We first we took an average of the 12 items available (half of the 24 learning objectives) for each condition (*massed* or *spaced*) and question instance (*1*, *2*, and *3*) for each student. We then assessed the difficulty of spacing over time with a two-way repeated-measures analysis of variance (ANOVA), using *condition* and *question instance* as within-subjects factors. Simple main effects were assessed with *t* tests as needed.

It should be noted that *question instance* indicates timing for the spaced condition, but *only* for the spaced condition. In the massed condition, three variants of the question were asked on the same quiz in a random order. In the spaced condition, the question was first asked in the quiz immediately following content instruction (similar to the massed condition question instances), *question instance 2* had a 1-quiz delay (2-weeks), and *question instance 3* had a 2-quiz delay (4-weeks).

III. Results

The ANOVA revealed significant main effects of *condition*, $F(1, 182) = 9.03$, $p = .003$, $\eta_p^2 = .05$, and *question instance*, $F(2, 181) = 24.67$, $p < .001$, $\eta_p^2 = .21$. These main effects were qualified by a significant *condition* × *question instance* interaction, $F(2, 181) = 21.77$, $p < .001$, $\eta_p^2 = .21$.

Estimated marginal means (Table 1) illustrate that performance was lowest on *question instance 1* in the *spaced* condition. Results from *t* tests indicated that performance on *question instance 1, spaced* was significantly lower than performance on *question instance 1*, *massed,* $t(182) = 6.03$, $p < .001$, Cohen's $d = .45$. The mean difference in performance was 6.7%. Performance on *question instance 2, spaced* was significantly lower than *question instance 2, massed*, $t(182) = 2.46$, $p = .015$, Cohen's $d = .18$, with a mean difference of 2.8%. There was no significant difference between *question instance 3* in the *massed* and *spaced conditions*, $t(182) = -1.31$, $p = .192$. In this third question instance, mean performance in the spaced condition was 1.4% higher than in the massed condition.

TABLE 1. ESTIMATED MARGINAL MEANS FOR THE CONDITION × QUESTION INSTANCE INTERACTION.

Condition	Question Instance	Mean (%)	Standard Error	95% Confidence Interval	
				Lower Bound	***Upper Bound***
Massed	1	80.5	1.0	78.6	82.4
	2	81.1	1.0	79.1	83.1
	3	80.3	1.0	78.2	82.3
Spaced	1	73.8	1.2	71.5	76.2
	2	78.3	1.1	76.1	80.4
	3	81.7	1.0	79.7	83.7

Note: In the massed condition, the question instances represent three algorithmic variants of the same question asked in a random order on the same quiz. In the spaced condition, the question instances refer to the question variants being asked in chronological order on three sequential quizzes. Each presented mean is the average student performance on the 12 items in each condition and question instance.

IV. Discussion

Similar to prior work [6], [12], we found that performance was significantly worse on spaced quiz questions than massed ones. Difficulty associated with spacing was evident on the first and second questions but not the third. Performance on the first spaced question instance was 6.7% lower than performance on the massed question instance. As hypothesized, this difference is larger than the 2.7% average difference. Performance on the second question was 2.8% lower, similar to the average difficulty, whereas performance on the final question was 1.4% higher.

Looking at the results in a different way, students performed better on objectives that had three similar questions on the quiz immediately following the weeks of content instruction, rather than objectives that had only one question. A possible explanation is that students learned *during* the quiz from seeing multiple instances of the same question. Mathematics classroom research has shown that a test is in fact a learning experience

[13]. It is possible that students learned through comparison [14], seeing multiple algorithmic variants of the same question in the massed condition. It is also possible that students found it easier to solve problems because the solution procedures were readily available in their short-term memory. It is not possible for us to observe the learning trajectory within a single quiz because we do not have a record of the order of the questions as they were asked to students. We are therefore not able to tell if students improved in performance "over time," performing higher on later questions, or if they compared between questions before generating a correct solution. We also cannot tell whether students went back and corrected earlier mistakes after noticing them on later questions. It is clear, however, that students performed better on the initial quiz when they saw three questions instead of one.

Observing that the difficulty was imposed at the beginning and middle of the spaced distribution has some implications for underlying learning mechanisms of spaced retrieval practice and desirable difficulty. The concept of desirable difficulty is that memory can be improved by making learning more difficult. Some researchers (including authors of the current publication) have proposed that the *delay* increases the difficulty due to the additional processes required in retrieval of stored knowledge. However, the temporal pattern of performance indicates lower *initial* learning. Still in line with desirable difficulty theories, these results shed important light on the temporal nature of the difficulty of retrieval practice.

A. Limitations

As always, replications are needed to determine whether this is a pattern of spaced retrieval practice in general or occurred due to some unique factor in this course or implementation. This study is limited to results from a single course in a single semester. However, this experiment was well-controlled, and there is no reason to believe this result is not indicative of a larger pattern.

Another limitation is that we did not look at whether the difficulty imposed by spacing was *desirable*, i.e., associated with superior retention of learning objectives at the end of the semester. We will test for that association in future work.

V. Conclusions

We presented a temporal analysis of the difficulty imposed by spacing in an engineering mathematics course. We found that spacing introduced the most difficulty right away, with students performing better on the spaced questions over time. These findings indicate that students may be learning in the spaced condition through feedback, perhaps in addition to strengthening recall processes.

The analyses presented here were possible because of the database created by the research support team. The ability to drill down into the data quickly and easily is a consequence of interdisciplinary collaboration between engineers and social science researchers.

References

[1] J. J. Pembridge and M. A. Verleger, "First-year math and physics courses and their role in predicting academic success in subsequent courses," in *Proceedings of the 120th ASEE Annual Conference and Exposition*, 2013.

[2] N. W. Klingbeil and A. Bourne, "A national model for engineering mathematics education: Longitudinal impact at wright state university," in *120th ASEE Annual Conference and Exposition*, 2013.

[3] D. G. Beanland, "Challenges and opportunities facing the education of engineers-Address to Victoria division of engineers Australia seg meeting." Melbourne, 2010.

[4] R. Suresh, "The relationship between barrier courses and persistence in engineering," *J. Coll. Student Retent. Res. Theory Pract.*, vol. 8, no. 2, pp. 215–239, 2006, doi: 10.2190/3QTU-6EEL-HQHF-XYF0.

[5] 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, doi: 10.1007/s10648-015-9349-8.

[6] K. B. Lyle, C. R. Bego, R. F. Hopkins, P. A. S. Ralston, and J. L. Hieb, "How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge," *Educ. Psychol. Rev.*, vol. 32, no. 1, pp. 277–295, 2019, doi: 10.1007/s10648-019-09489-x.

[7] 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, pp. 1432–1463, 2014, doi: 10.1037/a0037559.

[8] 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, pp. 354–380, 2006, doi: 10.1037/0033-2909.132.3.354.

[9] N. Kornell, "Optimising learning using flashcards: Spacing is more effective than cramming," *Appl. Cogn. Psychol.*, vol. 23, no. 9, pp. 1297–1317, 2009, doi: 10.1002/acp.1537.

[10] N. C. Soderstrom and R. A. Bjork, "Learning versus performance: An integrative review," *Perspect. Psychol. Sci.*, vol. 10, no. 2, pp. 176–199, 2015, doi: 10.1177/1745691615569000.

[11] E. L. Bjork and R. A. Bjork, "Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning," in *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2nd Editio., M. A. Gernbacher and J. Pomerantz, Eds. New York, NY: Worth, 2014, pp. 59–68.

[12] C. R. Bego, P. A. S. Ralston, K. B. Lyle, and J. C. Immekus, "Introducing desirable difficulty in engineering mathematics with spaced retrieval practice," in *Proceedings of the 128th ASEE Annual Conference and Exposition*, 2021, pp. 1–7.

[13] F. Avvisati and F. Borgonovi, "Learning mathematics problem solving through test practice: A randomized field experiment on a global scale," *Educ. Psychol. Rev.*, vol. 32, no. 3, pp. 791–814, 2020.

[14] J. R. Star *et al.*, "Learning from comparison in algebra," *Contemp. Educ. Psychol.*, vol. 40, pp. 41–54, 2015.