Work in Progress: Effect of Assessment Frequency on Long-Term Retention of Engineering Content

Dr. Prabhakar Venkateswaran, Milwaukee School of Engineering

Prabhakar Venkateswaran is an Assistant Professor of Mechanical Engineering at the Milwaukee School of Engineering. He received his Master's and PhD in Aerospace Engineering from the Georgia Institute of Technology, and his Bachelor of Science degrees in Mechanical and Aerospace Engineering from the University of Miami. His research and teaching interests are in the thermal sciences with a particular focus on combustion and gas turbines. In addition, he is also interested in investigating effective teaching and learning practices and issues affecting the persistence of under-represented minorities in STEM fields.

Michael D Cook, Milwaukee School of Engineering

Michael D. Cook is an assistant professor in the Mechanical Engineering Department at the Milwaukee School of Engineering (MSOE). He received the B.S. degree in electrical engineering from the University of Wisconsin-Madison, Madison, WI, USA, and the M.S. and Ph.D. degrees in mechanical engineering from Michigan Technological University, Houghton, MI, USA. His interests are in control system design and optimization of mixed-physics dynamic systems, with current research in power flow control with emphasis on the optimization and decentralized control of microgrids.

Dr. Michael Christopher Sevier, Milwaukee School of Engineering

Michael Sevier is currently an assistant professor of Mechanical Engineering at the Milwaukee School of Engineering. After finishing his doctorate degree at the University of California, Santa Barbara, Michael took a position at ATA Engineering where he worked as a structural analysis engineer for nine years. During this time, he both took and taught multiple professional courses and realized how many technically brilliant instructors struggled to convey information in a way that could be readily absorbed by the students. Now in his third year in academia Michael is researching how various teaching methods and study habits affect the absorption and long-term retention of class material in the hopes of best preparing students for their future as engineers.

Work in Progress: Effect of Assessment Frequency on Long-Term Retention of Engineering Content

Abstract

This work-in-progress paper presents some preliminary data from a study investigating the effect of assessment frequency on students' long-term retention of engineering course content. In this study, a variety of hypotheses designed to identify the factors that impact students' long-term retention of course content are tested. However, the primary hypothesis can be stated as follows: higher frequencies of assessment lead to improved long-term retention of course content. This hypothesis is investigated by implementing a different assessment frequency in two sections of the same course. In one section, students are given two midterms spaced roughly four weeks apart, while in the second section, students are given a twenty-five minute quiz roughly every two weeks. The students' long-term retention is then measured by their performance on the final exam, which is common to both sections. This study is conducted across three different core courses in the mechanical engineering curriculum; Principles of Thermodynamics II, Mechanics of Materials II, and Automatic Control Systems. The results from Thermodynamics II and Automatic Controls showed no statistically significant difference in student performance between the two assessment frequencies, while in Mechanics of Materials II, the section that was assessed more frequently performed significantly better on the final exam. However, the low sample sizes in each section decrease the statistical power of the study, making it challenging to arrive at meaningful conclusions.

However, the lessons learned from this study are being used to construct a study in the thermodynamics sequence, which will be conducted over multiple years to collate a sufficiently large sample size which will increase the power of the study. Furthermore, this study will seek to lengthen the time-scale over which student retention is being measured. In other words, in this study "long-term" is associated with the students' ability to retain content over the course of a 10-week quarter, while in the future study, students' content retention will be tracked over multiple terms over a complete sequence of courses. This will provide a more meaningful indicator of long-term content retention.

Introduction

Long-term retention of engineering content is an important requirement for students to be successful in an engineering curriculum and to have a successful career after graduating. Courses in a curriculum often build off each other and a weak foundation in a pre-requisite course can lead to students struggling and experiencing frustration in follow-on courses. Furthermore, increased long-term retention of engineering content can better prepare students to have successful and fulfilling careers after graduation, particularly in technical fields.

There are numerous studies in the literature that discuss a variety of strategies to increase student engagement in engineering courses, which in turn help them learn the material more effectively,

allow them to better persevere in an engineering curriculum, and consequently graduate in a timely manner [1]–[3]. In this study, we focus on the role of assessment in helping student learning because of the demonstrated connection between teaching, assessment, and learning [4]. The literature on educational assessment makes an important distinction between "assessment of learning", where students understanding of course content is being measured, and "assessment for learning" [5], where assessments are used as a tool to enhance learning [6]. In the book by Brown et al. [7], the authors summarize several studies demonstrating that frequent testing over a reading passage is a more effective method of enhancing learning compared to repeated reading of the same passage, despite many students claiming that they perceived repeated readings to be more effective compared to frequent assessment. Roediger and Karpicke [8] review various studies that demonstrate how assessment can be a powerful tool used to promote learning. In particular, they suggest that long-term learning and retention of course material can be promoted by designing frequent assessments that have certain key attributes. First, each assessment should be cumulative and include material from earlier parts of the course such that students are forced to continually revisit earlier parts of the course (retrieval practice) and in the process make important connections across the content of the course. However, assessments should be spaced such that students "forget" some of the earlier material forcing them to revisit and reconsolidate that material as they prepare. Studies found that if students were not given the opportunity to "forget" content between assessments, their long-term retention of content actually decreased, because the weakening and subsequent consolidation and strengthening of these neural connections was not occurring [7], [9]. Embedded in these conclusions is the recognition that increasing the number of assessments also provide opportunities for students to receive more frequent feedback, which can aid with their learning as well [5].

In this study the impact of assessment frequency on long-term retention of engineering content is investigated in three core courses in the mechanical engineering curriculum; Principles of Thermodynamics II, Mechanics of Materials II, and Automatic Controls. The outcomes from this study will help faculty design data-driven assessment strategies that promote long-term learning. Furthermore, the goal is to also help instill in students the importance of testing as a means to learn and assess learning, motivating them to undertake in frequent self-assessment, which is important once they leave school and cannot on rely on their instructors to provide them with assessments. As a result, this will help them develop into more effective life-long learners [6], which is embodied in a key ABET student outcome that students must demonstrate "an ability to acquire and apply new knowledge as needed, using appropriate learning strategies" [10].

Design of Study

Since this study focuses on assessment frequency on retention of course content, the central null hypothesis of this study was "*Assessment frequency will not affect students' retention of long-term retention*." This study was deployed in the following three core courses in the mechanical engineering curriculum during the Fall 2018 quarter; Mechanics of Materials II, Principles of Thermodynamics II, and Automatic Controls. For each course, one instructor was assigned to two sections, which allowed the instructor to deliver the course identically to both sections, but assess the two sections differently. In one section, students were administered two 50-minute exams, while students in the second section were administered five 25-minute quizzes. Students in both sections were given the same comprehensive final exam and their score on it was used as a measure of their long-term retention. All the quizzes and midterms were designed to be

cumulative as they continually assessed knowledge from earlier parts of the course. The quizzes took place approximately every two weeks, while the two midterms were scheduled for the 4th and 7th weeks of the quarter. Across all three courses, the average quiz grade, and average midterm grade constituted the same proportion of the students' grades in both sections of a course. In addition, the final exam score was given the same weighting for students in both sections as well. In other words, the average quiz score and midterm score constituted 45% of the final grade, and the final exam constituted 35-40% of the final grade. Some additional data was also collected to identify other controlling variables. These data included student GPA as well as student reported data to the following questions:

- How many hours of sleep did you get the night before and two nights before the assessment?
- How many other exams and/or quizzes did you have this week?

All data collection procedures complied with Institutional Review Board standards for the safekeeping of research participant information. Analysis of the data was conducted using the R statistical analysis software package [11]. The comparison of average common final results between the two sections was performed using a two-sided student *t*-test with a 5% significance level. The hypothesis test was conducted by stating the following null hypothesis, H_0 for a given course in this study:

H_0: The mean common final exam scores between the two sections will be the same i.e. $\mu_{exams} = \mu_{quizzes}$

where μ_{exams} is the average common final score from the section in which two exams were administered and $\mu_{quizzes}$ is the average common final score from the section in which five quizzes were administered. The alternative hypothesis, H_1 is expressed as:

H_0: The mean common final exam score between the two sections will not be the same i.e $\mu_{exams} \neq \mu_{quizzes}$

The *t*-test was selected due to the insufficient number of required to model the distribution as normal. The *p*-value is the probability that the difference in the of the common final results between two sections of a course, assuming that the null hypothesis is true, is due to random variations in the data. As a result, the lower the *p*-value, the less likely it is that the difference in the means is due to random variations, and the null hypothesis can be rejected. The 5% significance level means that if the *p*-value value for the difference in the means is less than 5%, the null hypothesis is rejected [12].

Results and Discussion

Table 1 summarizes the key data from the study, including the number of students in each section of the various courses involved in the study, the average major GPA of the students in each section, the average final exam average in each section, and the *p*-value from the *t*-test applied to the difference in the means of the final exam score between the two sections.

Table 1: Summary of key data for the courses involved in the study

Course	Section	No. of students	Average Major GPA	Assessment Frequency	Final Exam Average	p-value
Mech. Of Materials II	031/032	20	2.51	2 Midterms	64.2%	0
	041/042	16	2.91	5 Quizzes	79.0%	
Thermodynamics II	005	10	2.52	2 Midterms	76.3%	0.7097
	004	16	2.64	5 quizzes	74.0 %	
Automatic Controls	021/022	26	3.04	2 Midterms	73.0%	0.7514
	031/032	27	2.81	5 Quizzes	74.9%	

The results from the Thermodynamics II and Automatic Controls courses show that there was no statistically significant difference between the final exam scores of two sections, suggesting that assessment frequency did not have an impact on students' long-term retention of content. However, in Mechanics of Materials II, there was a statistically significant difference between the two groups of students, with the group being assessed more frequently displaying a stronger performance on the final exam. While the results for Mechanics of Materials II are promising, these results must be interpreted cautiously. As shown in the table, the number of students in each section is small, which makes the obtained data susceptible to outliers. More importantly, the small sample sizes lower the power of the statistical test, which implies that the probability of rejecting the null hypothesis when it is false is low. For this reason more detailed analysis taking into account other variables such as the amount of sleep and number of other assessments in the same week was not performed since this would have further segmented the data and lowered the power.

There are some additional factors worth noting when comparing the difference in outcomes across the three courses. In Mechanics of Materials II, the instructor chose to have the students complete "wrappers" after every assessment, which consisted of students having to re-work the exam or quiz after it was returned to them for a small percentage of their exam grade. These re-worked exams and quizzes were then graded on a binary scale based on accuracy. This was not done in Thermodynamics II and Automatic Controls. Compelling the students to actively review their assessments and make corrections may have had an important impact on their learning. An additional factor to consider is that the Mechanics of Materials II section that was assessed more frequently also had a higher average major GPA. As a result, the higher level of performance may also be attributed to higher-achieving students, if the GPA is taken to be a measure of academic achievement.

Conclusions

A study was designed and implemented to investigate the impact of assessment frequency on students' long-term retention of engineering course content. While there is some evidence that

frequent and cumulative assessments promote long-term retention, these results have to be viewed cautiously since the sample sizes in this study were quite small, which resulted in lower statistical power of the test. Furthermore, this study was implemented in courses that are in the middle of a sequence of courses as opposed to the introductory course in the sequence. As a result, students who did not develop a sound foundation in the introductory course may not benefit from either assessment strategy. As a result, a new study is being formulated which will be conducted over multiple years in a multi-course sequence. This will generate a larger sample size as well as extend the time-scale over which long-term retention is being measured compared to the ten weeks considered in this study. The larger samples size will also permit additional controlling factors to be considered such as student GPA, attending tutoring session, sleep habits, etc.

Acknowledgment

The authors would like to acknowledge the assistance of Kristin Shebesta for valuable input in designing the study. The authors would like to thank Michael Swedish and Kevin Hart for sharing data from their courses and for their input into the study design.

References

[1] M. Prince, "Does active learning work? A review of the research," *J. Eng. Educ.*, vol. 93, no. 3, pp. 223–231, 2004.
[2] R. M. Felder and R. Brent, *Teaching and learning STEM: A practical guide*. John Wiley & Sons, 2016.
[3] T. J. Kennedy and M. R. L. Odell, "Engaging students in STEM education.," *Sci. Educ. Int.*, vol. 25, no. 3, pp. 246–258, 2014.
[4] P. Black and D. Wiliam, "Assessment and Classroom Learning," *Assess. Educ. Princ. Policy Pract.*, vol. 5, no. 1, pp. 7–74, Mar. 1998, doi: 10.1080/0969595980050102.
[5] R. Hernández, "Does continuous assessment in higher education support student learning?," *High. Educ.*, vol. 64, no. 4, pp. 489–502, 2012.
[6] D. Boud and N. Falchikov, *Rethinking assessment in higher education: Learning for the longer term*. Routledge, 2007.
[7] P. C. Brown, H. L. Roediger III, and M. A. McDaniel, *Make it stick*. Harvard University Press, 2014.
[8] 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.
[9] L. R. A. BJORK, "On the symbiosis of remembering, forgetting, and learning," in *Successful remembering and successful forgetting*, Psychology Press, 2011, pp. 19–40.
[10] "Criteria for Accrediting Engineering Programs, 2020 – 2021." [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2020-2021/.
[11] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
[12] W. Navidi, *Statistics for Engineers and Scientists*, 3rd ed. McGraw-Hill, 2010.