2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #32340

Using Quizzes Effectively: Understanding the Effects of Quiz Timing on Student Motivation and Knowledge Retention

Major John Case, United States Military Academy

JOHN CASE is a U.S. Army Officer in his 12th year of active duty service and is an Assistant Professor in the Department of Systems Engineering at the United States Military Academy. He holds master's degrees in Operations Research from Virginia Tech and Engineering Management from the Missouri University of Science and Technology. He is in the Operations Research functional area of the Army and has research interests in operations research, data analytics, and education. He is certified in project management (PMP) by PMI.

Major Dereck Kennedy, United States Military Academy

Dereck Kennedy is an operations research systems analyst for the U.S Army and is currently an Assistant Professor in the Department of Systems Engineering at the United States Military Academy. He previously served in a variety of roles as an Engineer officer and is a certified as a project management professional (PMP) by PMI. He holds a M.S.E. degree in Industrial and Operations Engineering from the University of Michigan and a B.S. in Mechanical Engineering from the United States Military Academy.

Using Quizzes Effectively: Understanding the Effects of Quiz Timing on Student Motivation and Knowledge Retention

Abstract

For many years, teachers have leveraged the positive effects of quizzing and testing on knowledge retention in classrooms to improve student learning outcomes. Many aspects of quizzing have been investigated to further improve its use in the classroom, however, there have not been formal investigations to understand the impacts of quiz sequencing on student learning potential. In this study, we investigate the effects of frequent in-class quiz sequence, both pre-lecture and post-lecture quizzes, on student lesson preparation, participation in class, and knowledge retention. Utilizing a single factor experimental design and blocking by course, we gathered student performance data and student perception survey data to assess the impact of quiz sequence. Quiz sequence was found to have a significant effect on student motivation, in-class engagement, and knowledge retention. This paper presents findings and recommendations for teachers to consider in implementing frequent quizzing in their courses. The implication of this research is that a deliberate approach to the choice of a pre-lecture or post-lecture quiz will allow teachers to better meet their student learning objectives.

Introduction

Teachers, instructors, and professors at every level of education seek to improve student performance because a student that is doing well in a course is generally more participatory in class and motivated to invest the time and energy to do well on assignments and assessments. Additionally, effective learning and retention of fundamental concepts is essential to developing the depth of knowledge required when critically applying the material beyond the classroom [1]. Enabling effective learning in the classroom is not a simple task. Many factors must be considered when developing an approach, such as the student population being taught. In a single classroom, each student is unique in how they receive and process information based on factors such as preferred learning style, personal experience, competence in fundamental concepts, and individual motivation [1]. Such considerations necessitate instruction strategies that incorporate teaching tools and techniques capable of reaching a broad spectrum of student profiles. Fortunately, the use of teaching tools and techniques has been supplemented by research to both understand their impact and the best means of employment to increase student learning potential. Much of the research delves into the underlying factors that, when present or absent, cause teaching methods to be more or less effective in enabling student retention of key concepts and capacity for using these concepts in a meaningful way. One such factor that is consistently found

to be beneficial to long term retention is effortful retrieval of information through activities such as testing. So, should we just test students more? Many students would likely balk at the idea of incorporating more testing. This is largely due to the way quizzes, tests, and examinations are commonly used. Their application is often limited to one of assessment. Unfortunately, much of the disdain associated with these types of activities is based on the level of stress and energy involved. As such, the natural disposition of some students and teachers is to avoid the inclusion of more tests because of the negative associations [2]. Unfortunately, this avoidance causes many to dispense with a powerful enabler in helping students not only perform better on assessments but grasp the material more effectively. The key is proper incorporation during the learning process. Fortunately, research has added greater nuance to our understanding of how and to what extent effortful retrieval practices such as testing add value.

The "Testing Effect"

Retrieval practice in support of learning is often referred to as the "testing effect". The use and benefits of the "testing effect" are not new. A quote from the philosopher and statesman Francis Bacon in 1620 conveys the fundamental idea of the"testing effect" [3],

> If you read a piece of text through twenty times, you will not learn it by heart so easily as if you read it ten times while attempting to recite from time to time and consulting the text when your memory fails

In 1890, the philosopher and psychologist James attributed the benefit of testing, recitation, and other forms of effortful recall to the active participation required of the learner in such activities as opposed to the more passive role assumed when re-reading material or receiving a lecture [4]. Recent meta-analysis and individual research efforts have continued to exhort and expand these ideas. One such meta-analysis of relevant research findings since 1917 concluded that testing is "a powerful tool to enhance learning", among several other important findings [2]. Before delving into these findings and those of subsequent research, we provide the context and motivation for this study.

Study Motivation

This research effort was motivated by a desire to leverage the benefits associated with the "testing effect" in two different courses within the Systems Engineering Department at the United States Military Academy, Statistics for Engineers and Decision Analysis. Both courses are quantitative in nature and taught primarily to third year undergraduate engineering students. It is important in these courses, as with most others, that students prepare prior to class by studying an array of readings and sample problems. The expectation for lesson preparation is further emphasized by West Point's Thayer method. The Thayer method places a significant onus on students for their learning by requiring preparation to not merely be familiarization with the material but sufficient understanding to actively apply the material during classroom activities and discussion. It is often easy for instructors to discern if students have prepared for a lesson appropriately because, if they have not done so, students will remain in "receive mode", merely waiting for the material to be pitched to them. This passive behavior, of course, is not conducive to students actually learning the material and, likely, represents a wasting of the limited time the instructor and class have together to develop the depth of knowledge students require for critical thought, our ultimate goal.

We seek to maximize this valuable time by leveraging the known benefits of the "testing effect" to influence student lesson preparation, participation in class, and retention of the material. We first review available research about the "testing effect" to legitimize its inclusion in our teaching approach and determine the factors that must be considered in effective application.

"Testing effect" Benefits

In the meta-analysis by Karpicke and Roediger [2], several direct and indirect benefits associated with the "testing effect" were recognized. Indirect effects are those that are not attributed directly to the testing itself; rather, these effects are associated with the use of testing. Though some of them are somewhat intuitive upon consideration, their association with the "testing effect" could be incentive enough for most teachers to incorporate more testing.

Indirect Benefits

First, more frequent testing encourages studying through the duration of the course rather than during massed periods immediately before an exam. Additionally, student feedback on course and instructor surveys in these studies consistently reported that lesson preparation felt more structured and there was a greater incentive to invest more time and energy to this effort [4], [5], [6]. Narloch et al. believed that these results were closely tied to teacher observations of greater class engagement as well as reports by students that the material covered in class was clearer and more organized, likely due to their prior exposure to the material during preparation[5]. Such findings support the cognitive research concept that prior knowledge of a topic improves learning and retention of new knowledge on the same topic [7].

Another indirect benefit from more frequent testing comes from the feedback [2]. When done well, timely and effective feedback provides students an in-stride assessment of their understanding of the material and an added basis from which to focus future study efforts and questions during class. This likely enables an additional indirect benefit, a reduction in test anxiety during major graded events. Many attribute the reduction in anxiety, along with improved performance, to the added exposure to the testing environment and question types experienced during subsequent assessments [2], [8]. This finding is closely tied to the principle of transfer appropriate processing which suggests that practicing tests that require active retrieval fosters good performance on tests that require active retrieval [9]. Interestingly, this does not necessitate the questions used during early tests be mapped directly to those used on later assessments; rather, the benefit of testing some concepts from a set of material enables retention of related concepts in the same material [4].

Direct Benefits

Determining direct benefits of testing in improving long-term retention is the primary focus of most "testing effect" empirical research. Namely, how is student retention, when represented as performance on assessments, effected by the use of testing as a teaching tool? Many studies are naturally composed of a control group that is solely reliant on initial study and subsequent re-study of material while the experimental group is subject to testing with multiple factors of some variable. Roediger and Karpicke looked at laboratory and classroom studies and found that

the results indicated a clear trend of increased performance in both environments [2]. Though each yielded predominantly positive trends for using testing, difference in performance was more significant in laboratory experiments than those found in the classroom. This is likely due to the presence of many other confounding variables found in the natural classroom environment that are difficult to control. However, the classroom is the environment we operate in, and, as such, our review of the research focused primarily on classroom-based studies. Below are a few, of many similar, example studies.

Batsell et. al. explored the external validity of the "testing effect" in an undergraduate introductory psychology class [4]. The experiment consisted of an experimental class that took a daily quiz on assigned textbook material while the control class only studied assigned reading. The mean score for quizzed students was consistently over 10 percentage points above non-quizzed students, even when blocking for different question types. Student feedback supported feelings of greater comfort with the material.

Braun and Sellers used short daily quizzes at the beginning of class with the intent of motivating students to attend, prepare for, and participate in an undergraduate accounting class [6]. Findings showed lower failure rates in groups that used the daily quizzes as well as increased student motivation to keep up with readings and actively participate during class.

In short, research supports the use of testing to enable greater retention of material. Given this potential for positive impact, complementary research has explored the best means of employing the "testing effect" to maximize its potential value.

Considerations for using the "Testing Effect"

There are many different variables that must be considered when evaluating the impact of the "testing effect". Researchers often choose one or a few of these variables per study. Consequently, research continues to add to our more nuanced understanding of the best conditions and means for encouraging the necessary effortful recall to improve learning. Several variables of note include testing frequency, test length, the amount of points assigned to the testing event, question type, and feedback type and timing.

Testing frequency is largely determined by the intent of the study. Studies focused exclusively on the impact to retention often explore the spacing of quizzes throughout a term, weekly or monthly, whereas studies with an additional focus on preparation and participation in conjunction with retention predominately incorporate daily quizzes. The consistently positive results associated with the research using daily quizzes validated an initial assumption that a daily frequency was appropriate for this study's application [4], [5], [6].

Test length is important because it drives the time and effort required to create, administer, and provide feedback. If the intent is to integrate this tool into classroom instruction, short quizzes, no more than two to five questions, are often used as the primary vehicle [2]. This same logic was applied when designing this study.

Another consideration when using testing is the points allocated to the activity. Educators must find the balance between proper incentive for effortful participation and avoidance of counterproductive testing anxiety. Low-stakes quizzes, often no more than 5-10% of the course

grade, have been used to this purpose with positive results in encouraging meaningful effort while still emphasizing learning rather than assessment [2]. Khanna showed that the positive effect of testing can be negated if the points assigned to the activity cause students to experience undue testing anxiety [10]. This study made testing low-stakes by allocating only 5% of the total course grade to performance on daily quizzes, the amount normally used for participation points in both courses.

Question type is generally classified by whether questions require recognition or production of information. Recognition questions include types such as multiple choice or matching while production questions, such as short answer or fill in the blank, provide no reference and require the student to pull information exclusively from memory. Studies that compare the two have generally found that the added effort required for production questions yield better results during subsequent exams [2]. This does not imply that recognition questions are unproductive. On the contrary, Roed et al. [8] used short quizzes with multiple choice questions to aid retention of key concepts in a 6th grade social studies class. Mean course scores of quizzed students in this context where a letter grade higher than that of their non-quizzed peers. The key is crafting questions to have a "desirable difficulty" that requires students to think about and retrieve the information [2]. Instructors used a combination of recognition and production type questions including true/false, select all that apply, multiple choice, and fill in the blank. The type was based on the most appropriate means for eliciting effortful recall of the current material.

The quality and timing of feedback has important impact on the retention and learning potential associated with the testing activity. As mentioned previously, proper feedback can enable several indirect and direct benefits. Conversely, omitting feedback, significantly delaying feedback, or providing uninformative feedback can inhibit or prevent retention and testing benefits [2]. Feedback during this study was provided immediately and accompanied by in-class discussion in an attempt to maximize its value to students.

Available research provides a good baseline for effectively incorporating testing in the classroom as an enabler to the student learning process. This study seeks to leverage and add to these findings by exploring an additional variable, the sequencing of daily quizzes during a lesson.

Contribution to the Classroom and Current Research

Although there has been much research regarding the use and format of daily quizzes, no research could be found that examined the impact of quiz sequencing on student learning potential. Sequencing in this context refers to the decision to administer the quiz either pre- or post-lecture. In the reviewed research, the quiz timing during the lesson was pre-determined as a held-constant factor. We believe this choice of sequencing impacts how the testing activity influences student participation and performance. In this study, we seek to understand the best means of implementing in-class quizzes as a recall activity by establishing if there is a discernible advantage in the timing of these quizzes during a class period. Specifically, we explore the effect of quiz sequencing on student lesson preparation, participation in class, and material retention.

Methodology

Selecting the Response Variables

The response variable needs to measure retention of critical course concepts in the respective course. Therefore, we believe that grades in major cumulative graded events, like Written Partial Reviews (WPR), Term End Exams (TEE), and course projects are the best measurable response variables to measure retention. The reason retention is of interest is because the potential to build upon lower level concepts is higher if students are retaining the material already learned in previous lessons. Therefore, we will have three response variables: WPR grades, TEE grades, and project grades. We also desire to analyze the students' perspectives of the effect of quiz sequence on their learning, motivation, and engagement. Therefore, we analyze student survey responses gathered under the same conditions as the student performance data. The student survey responses were gathered at the end of the course and used to assess student personal perception of the effects of quiz sequencing.

Choosing Factors, Levels, and Ranges

In order to design our experiment, we organized the factors that we believe will affect the response variables by factors that we will control in the experiment, held constant factors, nuisance factors that we can control, and the nuisance factors that we cannot control. This breakdown is shown in Figure 1 in a cause and effect diagram.

The controllable design factor will be the quiz sequencing. The levels of this factor will be the beginning of class and the end of class. For the quiz that is given at the beginning of class, referred to as the pre-lecture quiz, the assessment will be administered prior to the lesson being taught to the students. The end of class quiz, referred to as the post-lecture quiz, will be administered after the lesson material is taught to the class. The choice of these two levels will be to evaluate the effect of assessments prior to formal instruction or after receiving instruction.

In order to isolate the effect of quiz sequence, we must attempt to hold constant other potential sources of variability. First, the quiz method must be uniform across the course sections. We chose to use a mix of question formats, limited to true or false, select all that apply, multiple choice, and fill in the blank. The literature on quizzing supports use of these question format and they work well for quantitative course content. This quizzing method also lends itself well to the goal of retention, as identified by Kang et al [11]. Second, we must hold the lesson material covered in each lesson constant between the two different sections in each course. Third, each instructor will keep individual teaching style consistent between the two different sections in each course. These factors, if allowed to vary, would confound the effect of quiz sequence in our experiment.

The nuisance factors that we will control in our experiment will be the known and unwanted sources of variability in grades. Therefore, we will block instructor and course, since we have two different instructors and two different courses. Because each instructor teaches one course, the effect of instructor and the effect of course are inseparable in the data. The block variable is referred to as "course" in the analysis sections that follow, however we understand that the effect of instructor is also included in this variable. Blocking by course will effectively isolate the

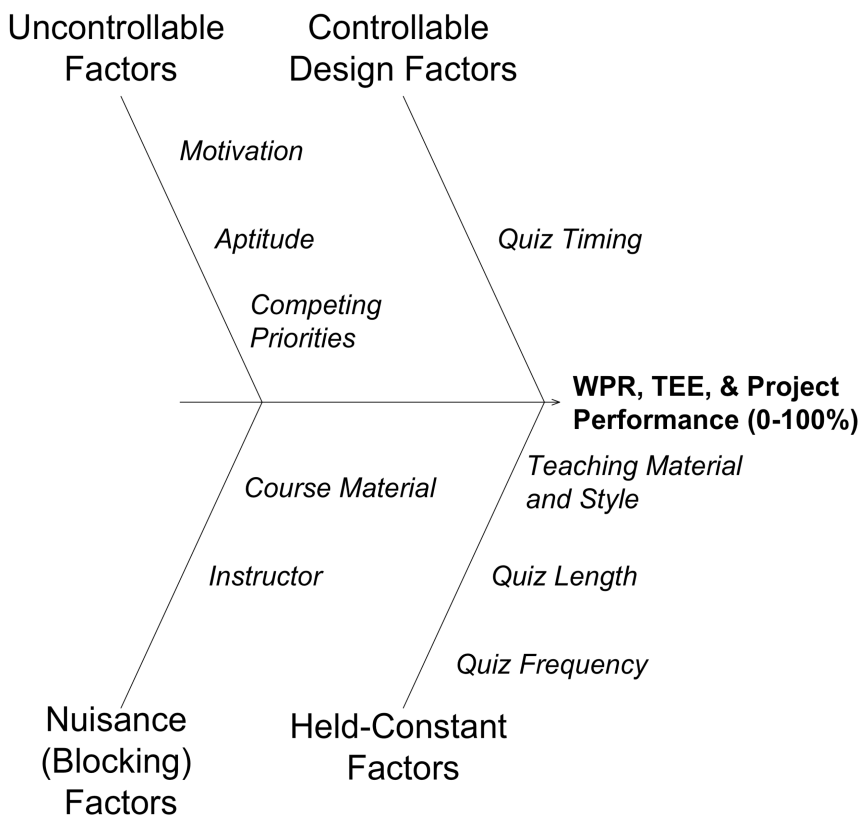

Figure 1: The Cause and Effect Diagram for the Quizzing Study

known and unwanted effect of different instructors and lesson material on students' grades.

The uncontrolled nuisance factors in this experiment will be the motivation of individual students and the variable nature of competing requirements for different students in different sections. This would be due to students' internal motivation, their interest in the course material, or the conflicts of being involved in different activities and other coursework that compete for their out of class study time. Because students are randomly allocated to course sections during registration, we believe that there will be no significant difference in students' aptitude or motivation when aggregated across different sections.

The Experimental Design

This is a single factor experimental design with a block. The experiment was replicated across 98 students, with equal numbers of students enrolled in the pre-lecture and post-lecture quiz sections. This experimental design should allow us to easily analyze and interpret the effect that quiz sequence has on grades in major cumulative events.

Data Sources

The data collected in this study came in two forms. The first was survey responses to a questionnaire regarding student perceptions of the quizzes and their effect on performance and

motivation. This data was analyzed to detect differences in survey responses between the pre-lecture and post-lecture quiz students. The second data source was student course performance grades on WPRs, TEEs, projects, and daily quizzes. This data was analyzed by fitting regression models to answer research questions regarding use of quizzing.

Survey Analysis

Survey Responses and Student Motivation

Each section of students was given a survey at the end of the semester to capture their perceptions of how the quiz timing affected their motivation and knowledge retention. The goal was to isolate how students' perceived their motivation in class and outside of class.

Free Text Response Analysis

The first block of questions and response types used on the survey was, "Based upon the fact that you were repeatedly quizzed at the beginning/end of class, how did the timing of the quiz affect:"

- Your preparation for class? [Free text response]
- Your in-class participation and note taking? [Free text response]
- Your motivation in or out of class? [Free text response]

We classified the free text responses to these three questions into three respective binary variables, marking a "1" if the response indicated there was an effect on their default behavior and motivation or a "0" if there was no effect. The binary coded variables capture whether the timing of the quizzes affect a student's motivation to prepare for class (outside of class) or their motivation to participate and take notes during class. In analyzing responses to "How did the timing of quizzes affect your preparation for class?," we found that although 44.7% of pre-lecture quiz students had increased motivation to prepare for class while only 32.6% of post-lecture quiz students had increased motivation to prepare for class, this was not a statistically significant difference in proportions. However, when this question was re-asked in the form "How did the timing of quizzes affect your motivation in or out of class?," the difference in pre-lecture and post-lecture quiz students was significantly different at the 0.02 level. 40.4% of pre-lecture quiz students had increased motivation to prepare for class, while only 15.2% of post-lecture quiz students had increased motivation to prepare outside of class time.

In analyzing responses to "How did the timing of quizzes affect your in-class participation and note taking?," we found that only 38.3% of pre-lecture quiz students had increased motivation to participate and take notes in class while 63.0% of post-lecture quiz students had increased motivation in class. This result was statistically significant at the 0.03 level. When this question was re-asked in the form "How did the timing of quizzes affect your motivation in or out of class?," the difference in pre-lecture and post-lecture quiz students was significantly different at the 0.0001 level. Only 6.4% of pre-lecture quiz students had increased motivation to participate and take notes during class, while 43.5% of post-lecture quiz students had increased motivation to participate and take notes during class.

This analysis suggests that the timing of daily quizzes has a significant effect on student motivation and engagement. If motivation to prepare outside of class is deemed more important than in-class participation, then the pre-lecture quiz is preferred. If in-class engagement and participation from students is deemed more important, then the post-lecture quiz should be used.

Likert Scale Response Analysis

The second block of questions and response types used on the survey was, "Rate each question response that follows on a Likert scale (Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree):"

- The quizzes improved my ability to learn and retain critical course concepts.
- The quizzes negatively affected my ability to learn or retain critical course concepts.
- I liked the consistent frequency of quizzes.
- I liked the timing of the quizzes.
- The amount of points associated with each quiz was appropriate.
- My effort for lesson preparation would increase if the quizzes were worth more of my overall grade.
- The amount of time each quiz consumed was appropriate.
- The level of quiz complexity was appropriate to their timing before/after class.

To analyze responses to the Likert scale questions, we compared the distribution of responses of the pre-lecture quiz sections to the post-lecture quiz sections. The statistical analysis technique used was the Mann-Whitney-Wilcoxon non-parametric test for independence of distributions. This method is useful because the hypothesized distribution does not have to be Gaussian. In this test, the null hypothesis assumes that the two sample distributions being compared are generated from the same underlying distribution. If there is significant evidence that the samples are drawn from different distributions, then the null hypothesis is rejected.

The questions that had no statistically significant difference in response distribution are shown in Figure 2. In the first three questions in Figure 2, student responses indicated that the quiz complexity, time, and points allocated to each quiz was appropriate. This was useful feedback, informing us that how we structured and allocated points to the quizzes was seen as fair by the students. This also confirmed that our efforts held the quiz frequency, length, point allocation, question types, and immediate feedback mechanism constant across class sections, as was intended in our experimental design. The last question shown in Figure 2 showed that students in both the pre-lecture and post-lecture quiz sections generally agreed that their lesson preparation would have increased if the quizzes were worth more points. This is an important consideration for courses in which out-of-class preparation is very important, as seen in history or humanities courses where a significant amount of reading outside of class is required.

The questions that yielded statistically significant differences in the pre-lecture and post-lecture quiz sections' response distributions are shown in Figure 3. The post-lecture quiz section students

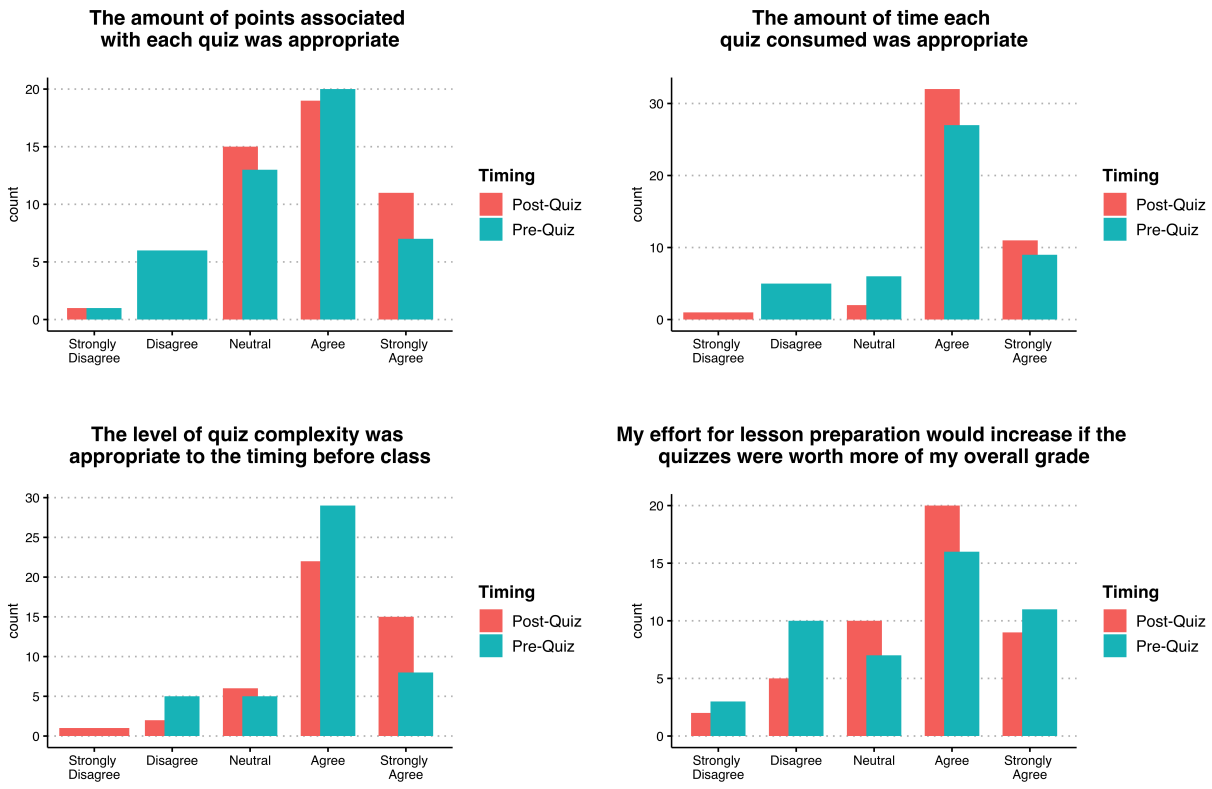

Figure 2: Responses without Significant Differences Between Pre-Lecture and Post-Lecture Quiz Sections

were more likely to respond affirmatively to the question “the quizzes improved my ability to retain critical course concepts.” This was an interesting result because we believed prior to executing the study that the pre-lecture quiz students would work harder outside of class, thus retaining more knowledge due to the effortful nature of this learning method. However, at least according to student self-perception, post-lecture quiz students were more likely to believe that the quiz timing improved their ability to retain critical course concepts. This indicates that in the very least, student confidence can be boosted by post-lesson quizzing.

While the pre-lecture and post-lecture quiz sections were shown to have significantly different response distributions to the question “the quizzes negatively affected my ability to learn or retain critical course concepts,” students in each section were approximately equally likely to respond negatively to this question. This was an affirmation from students that quizzing, regardless of timing, was helpful to learn and retain information.

Students in the post-lecture quiz sections were more likely to respond affirmatively to the question “I liked the consistent frequency of the quizzes” than students in the pre-lecture quiz section. This was likely due to the pre-lecture quiz student perception that earning points on the pre-lecture quizzes was more difficult. It is the authors’ impression that pre-lecture quiz students were believed to view the quizzes as frequent additional stress and requiring more time commitment outside of class.

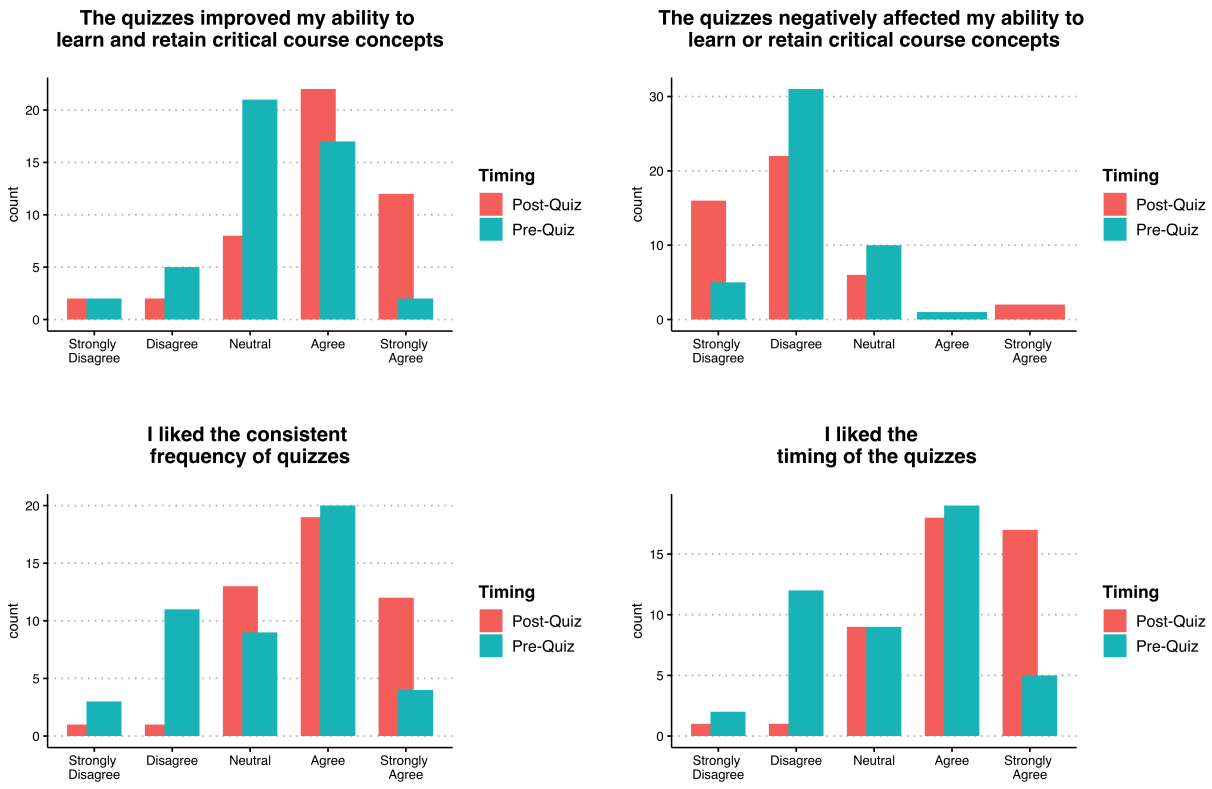

Figure 3: Responses with Significant Differences Between Pre-Lecture and Post-Lecture Quiz Sections

Finally, students in the post-lecture quiz sections were more likely to respond affirmatively to the question “I liked the timing of the quizzes.” This was again perceived to be due to the perception that students believed it took less effort to earn points on the post-lecture quiz. It was interesting to see that students in the pre-lecture quiz sections were more likely than not to respond affirmatively to this question, even when they knew that the post-lecture quiz was being implemented in other sections. One explanation for this is that students in both sections became accustomed to the timing of the quizzes and the effort required to do well and earn points. Students were observed to adapt to consistent requirements, whether they involved the pre or post-lecture quiz.

The biggest disagreement between pre and post-lecture quiz students was “I liked the timing”, where pre-lecture quiz students had 21.4% of students “disagree” while zero of the post-lecture quiz students “disagreed.” This result supports the theory that the post-lecture quiz helps boost the confidence of students. It also indicates that the pre-lecture quiz requires more effort and time commitment for the student to earn points, which students don’t like, but the literature indicates it is beneficial for learning and knowledge retention. The authors believe it also builds good habits in students that will hopefully continue with other coursework.

Student Performance Analysis

Student Performance and Knowledge Retention

We sought to investigate whether the timing of quizzes or quiz performance had any effect on major graded events like the WPR, project, or TEE performance across the two courses. We also wanted to address whether performance on the quizzes matters for knowledge retention or if quizzing itself was the benefit to student knowledge retention and learning outcomes. If the timing of quizzes was important, then we may observe an interaction effect between quiz timing and quiz performance in determining major graded event performance.

We know that certain variables that are not the subject of the study have influence on student performance, namely the students' incoming grade point averages, which is a measure of their past performance, motivation, discipline, and study habits. Additionally, the course itself will impact performance, since grade distributions are different across difference courses. These nuisance variables, GPA and course, can be included in the model building process in order to isolate or account for their effect on the variability of graded event scores. Therefore, our approach was to investigate our research questions by fitting multiple linear regression models to WPR, project, and TEE percentage scores, with independent variables of student GPA, quiz timing, and course.

Table 1: WPR and Project Regression Results

	Dependent variable:	
	WPR Scores	Project Scores
	(1)	(2)
Grade Point Average	14.889***	6.530***
	(2.178)	(1.142)
Stats Course	-12.378***	3.039**
	(2.840)	(1.489)
Quiz Score	0.212	-0.038
	(0.133)	(0.070)
Pre-Quiz	-10.371	-4.525
	(14.070)	(7.376)
Score:Pre-Quiz	0.087	0.044
	(0.215)	(0.113)
Constant	25.058***	71.527***
	(9.513)	(4.987)
Observations	98	98
Adjusted R^2	0.494	0.262
Residual Std. Error (df = 92)	10.919	5.724
F Statistic (df = 5; 92)	19.925***	7.880***
Note:	*p<0.1; **p<0.05; ***p<0.01	

When independently modeling WPR and project in terms of GPA, the course, student quiz performance, quiz timing, and the interaction between quiz performance and quiz timing, only GPA and course were significant predictors of WPR and project performance. The regression output for the WPR and project models is shown in Table 1. The model adjusted r-squared, a measure of the variability in the response variable accounted for by the independent variables, was below 50% for both the WPR and project models. This indicates a lack of fit for the regression model, meaning that the model does not adequately account for the variability in WPR and project scores. To analyze the effect of the significant variable, course, the regression output shows that WPR scores were 12.4% lower while project scores were 3.0% higher for students in the Statistics for Engineers course versus Decision Analysis. The quiz timing variable was not found to have a significant effect on WPR or project scores, even though WPR scores were 10.4% lower and project scores were 4.5% lower when the pre-lecture quiz was used.

When fitting the same independent variables to TEE score and overall course performance, we saw different results, shown in Table 2. While GPA and course were again significant predictors of the response variable, now quiz percentage was also significant and the adjusted r-squared was 58.7% and 69.2% in the TEE and course performance models respectively.

Table 2: TEE and Course Grade Regression Results

	Dependent variable:	
	TEE Scores	Course Scores
	(1)	(2)
Grade Point Average	10.084***	8.008***
	(1.305)	(0.697)
Stats Course	-3.655**	-3.220***
	(1.703)	(0.909)
Quiz Score	0.423***	0.085**
	(0.080)	(0.043)
Pre-Quiz	12.981	-6.426
	(8.433)	(4.502)
Quiz Score:Pre-Quiz	-0.174	0.094
	(0.129)	(0.069)
Constant	26.315***	59.312***
	(5.702)	(3.044)
Observations	98	98
Adjusted R^2	0.587	0.692
Residual Std. Error (df = 92)	6.544	3.493
F Statistic (df = 5; 92)	28.522***	44.577***
Note:	*p<0.1; **p<0.05; ***p<0.01	

This suggests that student quiz performance was a significant factor in determining TEE score, which is the cumulative graded event that is seen to best measure student knowledge retention and

measure student outcomes for the course. The effect size of quiz score was small, only 0.366 points, but still a statistically significant factor in determining TEE score. Because quiz score was also significant in the course performance model, we believe that it may be a demonstration of the importance of motivation and positive study habits, which would contribute to the variability in quiz scores. It is important to note that the relationships identified here are indicators of correlation, rather than causation. It is again noteworthy that the effect of motivation is confounded with the quizzing performance scores.

We also sought to understand if the effect of the timing of quizzes is dependent on course subject matter. If this is the case, then we should be able to observe a significant interaction effect between course and quiz timing. We chose to fit models with the course and quiz timing interaction, along with GPA, course, and quiz performance. When fitting models to our major graded events, we observed a significant interaction effect when TEE was the dependent variable. The regression output for this model is shown in Table 3.

Table 3: TEE Grade Regression Results

	Dependent variable:
	TEE Scores
Grade Point Average	9.834***
	(1.285)
Stats Course	-0.906
	(2.257)
Quiz Score	0.366***
	(0.070)
DA Course:Pre-Quiz	3.370**
	(1.678)
Stats Course:Pre-Quiz	-2.441
	(2.558)
Constant	30.221***
	(4.834)
Observations	98
Adjusted R^2	0.596
Residual Std. Error	6.470 (df = 92)
F Statistic	29.606*** (df = 5; 92)
Note:	*p<0.1; **p<0.05; ***p<0.01

The interaction effect of course and quiz timing variables suggests that the optimal timing of the quizzes is material-dependent and instructor-dependent. The interaction plot is shown in Figure 4, where the mean of TEE scores for students in each course and quiz timing section was computed. This plot shows that the effect of quiz timing on TEE scores was course dependent. Statistics for Engineers students were worse off on the TEE if they had the pre-lecture quiz, while Decision Analysis pre-lecture quiz students were better off on the TEE.

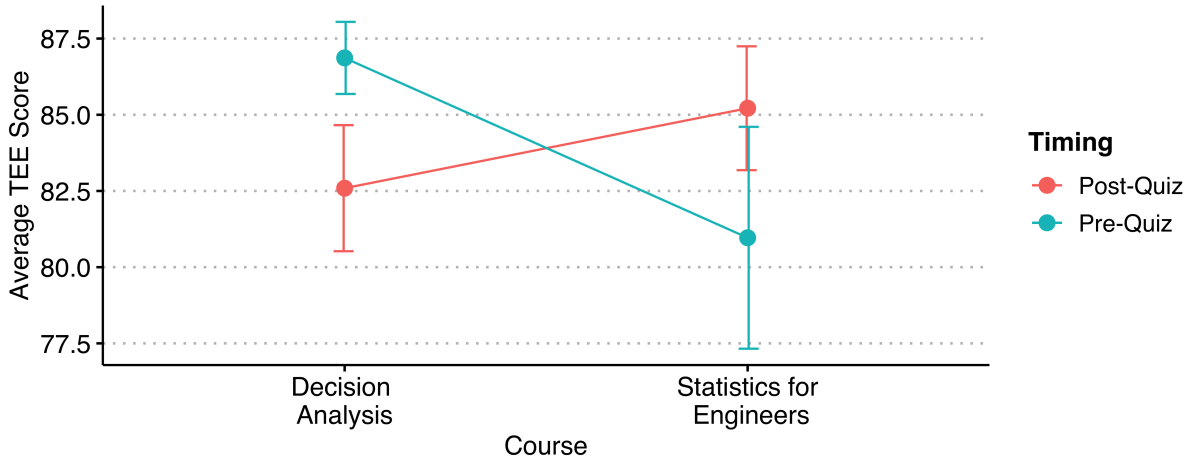

Figure 4: Effect on TEE Scores of Interaction of Course and Quiz Timing

We believe this may be because the Statistics for Engineers text is dense and more difficult for students to read and understand outside of class, while the Decision Analysis text is more readily absorbed. This result suggests that the choice for a pre-lecture or post-lecture quiz should consider the subject matter itself, as well as the density of the course text and other course resources.

Lower Performing Student Analysis

In many situations, maximizing the outcomes for the lowest performing students may be a focal point for instructors.

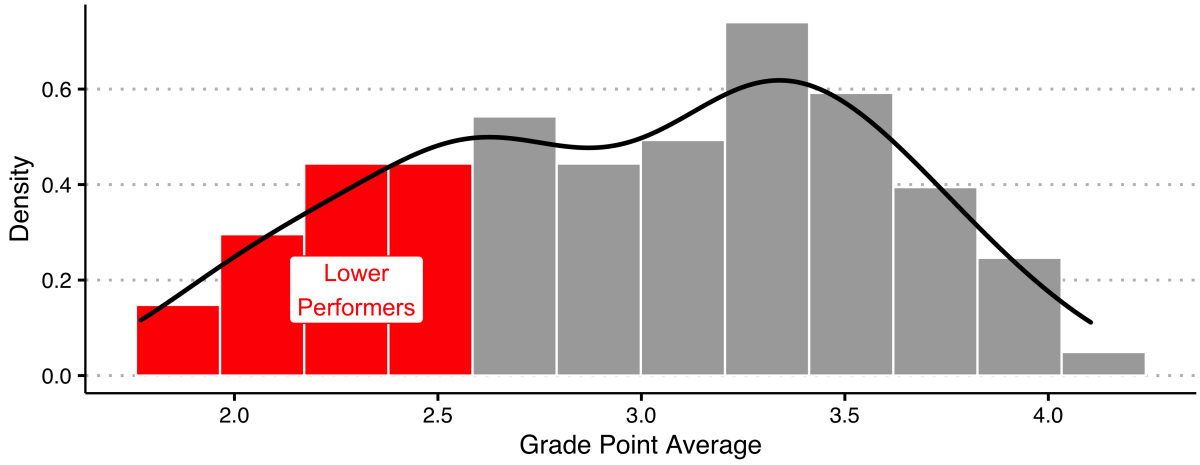

Figure 5: Effect on TEE Scores of Interaction of Course and Quiz Timing

To analyze the effect of quiz sequencing on lower performing students, the sample was split

according to figure 5, with a GPA cutoff of 2.5. This corresponds to a ”C” average and is a logical cut point for lower performing students. The sample was split into subsets of the lower performance students, yielding 22 students with GPAs below 2.5, and higher performing students, yielding 76 students with GPAs above 2.5. Then the same regression model was fit to each subset using TEE score as the response and independent variables of GPA, course, quiz timing, and the interaction between the course and quiz timing. The lower performing student model (model (1) in Table 5) shows that only the interaction of course and quiz timing was significant. Students that took the pre-lecture quiz in the statistics course scored an average of 22.4 % lower than their poor performing counterparts who either took the post-lecture quiz in statistics or took either pre- or post-quiz in the decision analysis class.

Table 4: TEE Grade Regression Results for Lower (1) vs Higher (2) GPAs

	Dependent variable:	
	TEE Scores	
	(1)	(2)
Grade Point Average	16.578	11.291***
	(9.553)	(1.933)
Stats Course	7.746	1.995
	(4.524)	(2.858)
Pre-Quiz	6.411	-0.111
	(5.360)	(1.869)
Stats Course:Pre-Quiz	-22.361***	-1.581
	(7.419)	(3.793)
Constant	36.593	50.850***
	(21.593)	(6.163)
Observations	22	76
Adjusted R^2	0.327	0.302
Residual Std. Error	7.995 (df = 17)	6.938 (df = 71)
F Statistic	3.546** (df = 4; 17)	9.131*** (df = 4; 71)
Note:		*p<0.1; **p<0.05; ***p<0.01

As previously discussed, the statistics text and material is more dense and difficult for students to grasp compared to decision analysis. The regression results indicate that lower performing students were adversely affected by the pre-lecture quiz in the more technical course, lending support for structuring courses so that lower performing students have a better chance at success. The implication of this result is that when a course has more difficult material or a dense text, avoidance of a pre-lecture quiz for lower performing students is prudent. This approach would suggest that students should either be sectioned according to past performance (GPA) or that no pre-quizzes should be given in dense-material courses. To verify that this approach would not adversely impact the learning experience of higher performing students, the same model was fit to the subset of students with GPAs exceeding 2.5. The only significant variable in this model (model (2) in Table 5) was GPA, indicating that quiz sequence had no effect on higher performing

students. This result lends support to structuring quiz sequence based on the expected effect on lower performing students. The pre-lecture quiz is more detrimental to these students' performance in courses when the material and text is relatively dense.

Student Performance Analysis Conclusions

There were multiple broader observations that stood out in the course of our regression model analysis. Student GPA was included in each regression model because it was presumed to be a significant predictor of performance on graded events. We saw that GPA is significant in each model and has the largest effect size in each model. GPA is a complex variable, because it captures the effects of student motivation, work ethic, discipline, and study habits and is at least partially confounded with the effects of quiz timing and quiz performance on motivation. Additionally, the interaction of quiz timing and quiz percentage was not significant in any models, suggesting that the effect of quiz timing on knowledge retention does not depend on how well you do on quizzes. Finally, there are multiple possible conclusions to draw from the fact that the best models were the TEE and course average models, which account for 57.1% and 65.4% of the variability in performance respectively. The first is that student motivation is likely the primary driver of performance over the entirety of the course, which would be captured the the overall course grade and to a somewhat lesser degree in the TEE grade. Secondly, if consistent quiz timing does indeed impact student motivation, then the course and TEE grade are the primary places we should observe this phenomena. Unfortunately, because inherent baseline student motivation is not separable from motivation due to quiz timing, we are unable to assign causality to quiz timing motivation.

Instructor Observations

Beyond the results of the experiment, instructor observations provided valuable insight when incorporating pre- or post-lecture quizzes, several that are difficult to capture through survey and performance data. The pre-lecture quiz proved useful in allowing instructors to quickly identify topics and concepts with which students were struggling, sometimes being different across sections. This allowed instructors to immediately re-assess what should be prioritized during subsequent lesson discussion and practice rather than waiting for feedback after class or through follow-on assignments. Pre-lecture quizzes also provided a good indication of student lesson preparation. Often times, this observation was not based on quiz performance; instead, students that prepared were more likely to ask questions or make comments that showed familiarity with the material. These questions and comments often provided a natural transition to and context for the lesson's learning objectives.

This experiment did not focus on the effects of using different types of questions, but instructors did notice that different material was better addressed with certain question types. The key consideration made during question formulation was deciding the best means of achieving a "desirable difficulty". This determination was largely dependent on the specific concepts being covered at that time. As such, it is recommended not to limit questions to a specific type.

Another observation was based on instructor preference for the quiz sequencing. Instructors that favor methodically walking through lesson objectives and exercises expressed preference for the pre-lecture quiz because of the insights it provides for prioritizing discussion. Instructors that

favor emphasizing the student's responsibility in the learning process conveyed a preference for the post-lecture quiz as a validation of student engagement. Preferences such as these further support the conclusion that the choice of using a pre or post-lecture quiz is largely contingent on the instructor style, the material being covered, and the desired effect for using the quiz.

Findings and Recommendations

Summary of Findings

There are multiple important findings from this study. As previously stated, these findings are specific to the two engineering courses observed and the more homogeneous student population found at the US Military Academy. However, instructors of any discipline, in any context, can benefit from these findings as considerations for how they observe and assess the effects of tools they employ in the classroom.

1. The timing of daily quizzes has a significant effect on student motivation and engagement. If motivation to prepare outside of class is deemed more important than in-class participation, then the pre-lesson quiz is preferred. If in-class engagement and participation from students is deemed more important, then the post-lesson quiz should be used.
2. Lower performing students were adversely affected by pre-lecture quizzes, while higher performing students were not significantly effected by quiz sequence. If a maximin optimization strategy for student performance is desired, using a post-lecture quiz is best.
3. The quiz structure, length, allocation of points, question types, and feedback mechanism used in the quizzes must be seen as appropriate to the material and judged as fair by the students in order to leverage maximum motivation and engagement in the class.
4. Post-lesson quiz students were more likely to believe that the quiz timing improved their ability to retain critical course concepts, suggesting that student confidence can be boosted by post-lesson quizzing.
5. Pre-lesson quizzes may be observed by students as more stressful than post-lesson quizzes.
6. Students adapt to consistent and clearly communicated quizzing requirements, regardless of the timing of quizzes.
7. Quiz performance is a statistically significant predictor of knowledge retention on term end exams, indicating that students who are motivated and engaged, either outside of class or in class, will demonstrate improved learning outcomes.
8. The effect of a pre-lesson or post-lesson quiz on graded event performance depends on the subject matter, course materials, and instructor's teaching style.

Recommendations to Teachers

1. Use quizzing as a learning tool to improve knowledge retention in coursework.
2. Use frequent or daily quizzing to improve student preparation and engagement in class.
3. In choosing pre-lecture or post-lecture quiz timing, teachers should decide which is more

important: student preparation (pre-lecture quiz preferred) or student engagement (post-lecture quiz preferred). If maximizing lower performing student outcomes is desirable, then a post-lecture quiz should be considered.

4. Considerations for pre-lecture or post-lecture quizzing implementation includes:
    - The quiz structure, length, point allocation, question type, and feedback mechanism.
    - The course material to be taught.
    - The text and course resource density or difficulty.
    - Student academic maturity.
    - Instructor teaching style.
    - Mitigate the negative effects of test anxiety by making the quiz low-stakes.
5. Teachers should calibrate their quiz implementation strategy through use of a student feedback mechanism (surveys) in order to ensure that the intended effects of quizzing are being felt by students. Note: Students may not be initially ”happy” with your chosen quizzing style, but over time they will adjust and become accustomed to the consistency of quizzing.
6. Whatever quizzing strategy is selected, teachers should have confidence that quizzing is a proven technique to improve student learning outcomes.

Acknowledgment

The authors would like to thank the many people who made this research possible. Dr. Finn, Dr. Evans, and LTC Bruhl were instrumental in the authors' development by providing excellent instruction and discussion in the Master Teacher Program, helping to inspire the authors to explore how to improve student motivation and knowledge retention. Most importantly, we thank the cadets involved in this study for their effort, candid feedback, and continued desire to improve their knowledge, skills, and abilities.

References

[1] Peter C Brown, Henry L Roediger III, and McDaniel Mark A. *Make it Stick: The science of successful learning*. Belknap Press of Harvard University Press, 2014.

[2] Henry L Roediger III and Jeffrey D Karpicke. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3):181–210, 2006.

[3] Jeffrey D Karpicke and Henry L Roediger III. The critical importance of retrieval for learning. *Science*, 319(5865):966–968, 2008.

[4] Robert W Batsell Jr, Jennifer L Perry, Elizabeth Hanley, and Autumn B Hostetter. Ecological validity of the testing effect: The use of daily quizzes in introductory psychology. *Teaching of Psychology*, 44 (1):18–23, 2017.

[5] Rodger Narloch, Calvin P Garbin, and Kimberly D Turnage. Benefits of prelecture quizzes. *Teaching of Psychology*, 33(2):109–112, 2006.

[6] Karen W Braun and R Drew Sellers. Using a "daily motivational quiz" to increase student preparation, attendance, and participation. *Issues in Accounting Education*, 27(1):267–279, 2012.

[7] John R Anderson. Effects of prior knowledge on memory for new information. *Memory Cognition*, 9:237–246, 1981.

[8] Henry L Roediger III, Pooja K Agarwal, Mark A McDaniel, and Kathleen B McDermott. Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17(4):382–395, 2011.

[9] J D Bransford, J J Franks, C D Morris, and B S Stein. Some general constraints on learning and memory research. In L S Cermak and F I M Craik, editors, *Levels of processing in human memory*, pages 331–354. Hillsdale, NJ: Erlbaum, 1979.

[10] Maya M Khanna. Ungraded pop quizzes: Test-enhanced learning without all the anxiety. *Teaching in Psychology*, 42(2):174–178, 2015.

[11] Sean H K Kang, Kathleen B McDermott, and Henry L Roediger III. Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5):528–558, 2007.