An analysis of using online testing modules in engineering economics

by

Vrishtee Rane

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial Engineering

Program of Study Committee:
Cameron MacKenzie, Major Professor
Michael Dorneich
Holly Bender

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

DEDICATION

My work is dedicated to my mother, Disha Rane, who has been a constant support throughout my life.

“Mumma, thank you.”

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my major Professor, Dr. Cameron MacKenzie for his guidance and for the time he invested in me, and my committee members, Dr. Dorneich, and Dr. Bender for their support throughout the course of this research.

In addition, I would also like to thank my friends for constantly motivating me, my colleagues, the department faculty and staff for making my time at Iowa State University a wonderful adventure. I want to also offer my appreciation to my students who were willing to participate in my surveys and observations, without whom, this thesis would not have been possible.

ABSTRACT

Undergraduate students often find it difficult to manage curriculum. Engineering courses often require student to solve assignments and take in-class exams throughout the semester that include solving complex practical applications which can be challenging, one of the reasons being the time constraints. Additionally, the use of computers to solve the real-world problems during in-class exams is demanding due to its set-up cost and accommodation of large number of students at the same time. This thesis describes the study of a large engineering class using a non-conventional testing method. Student performance was evaluated using online testing modules with a stringent passing criterion and the tests could be taken multiple times. The questions for each testing attempt was pulled from a huge database so that students received a new question every time. Student survey results indicated that most favored the online testing assessment method. Our results show that students learned from their mistakes and their performance improved by taking the test multiple times. We compared the performance of students on the traditional assessment who took the course in the previous semester with the performance of student on the online testing module. Our analysis shows that the students earned better grades using the online tests than the students in their in-class exams. The online assessment method could be useful in the large engineering courses that are focused on formula-based curriculum.

CHAPTER 1. INTRODUCTION

Since the advent of computers, a revolution has been seen, starting from the military to various industries. The 20th century witnessed the acceptance of the Internet in almost all life sectors including education. With the growing enrollment of students, in-class and for distant education, there is a need to upgrade the traditional methods of teaching and learning to reach all the students effectively and help students learn efficiently. While, the Internet is being used to deliver lectures for the distant learning students, refer the prior work in multiple research areas, submit the homework and communicate with students across the world, we still need to facilitate the students with the benefits of the internet.

To experiment with the traditional method of testing, the in-class exams, assignments, and homework, we utilized the Internet to develop and deliver online testing modules for the engineering economics course offered by the Department of Industrial and Manufacturing Systems Engineering. The online testing modules could be taken multiple times until the students passed the test. Every attempt on the testing module pulled a set of new questions from a large question bank such that no student received the same question more than once. The passing criteria for these testing modules were stringent. The aim of this experiment was to understand if the students performed better in the online tests than in the traditional exams, and to analyze if student learned from their mistakes and improved over the multiple attempts that they took on the online testing modules. The memorandum to Institutional Review Board (IRB) approved the exemption form for this experiment, which is attached in the Appendix section.

Chapter 2 describes the structure of the online testing modules and its grading scale. This chapter gives the insight of preferences of students and their opinions on the online testing assessment method. It depicts the performance of students on the online testing module, effect on the performance based on the major and course delivery, and improvements of students with multiple attempts of online testing.

Chapter 3 compares the performance of students assessed via the online testing modules and via traditional assignments and in-class exams. First, a comparison is made of the same set of students who had in-class exams and online final. This chapter also analyzes the performance of students in the in-class exams in spring 2017 with the performance of students on the online testing modules in fall 2017. The student performance on semester grades and specific questions between the two classes are analyzed. Finally, this thesis provides information about recommendations of students regarding the traditional assessment methods and non-conventional methods for assessing the academic courses.

CHAPTER 2. **USING ONLINE TESTING MODULES TO ASSESS STUDENT PERFORMANCE IN LARGE-ENROLLMENT ENGINEERING CLASSES**

Introduction

Engineering colleges in North America have experienced a consistent rise in student enrollment since 2007 (Yoder, 2012). The three largest university programs in Canada are electrical engineering, mechanical engineering, and computer engineering (Anderson & Gilbride, 2003). With such large engineering classes, students may have limited access to faculty and teaching assistants (Adrian, 2010), and classes may be taught in big lecture halls with little personalization for individual student preferences (Gomes & Mendes, 2007). Large student enrollment in engineering classes and the rise of technology raise questions of how best to assess student performance.

In many engineering courses, students are graded based on weekly homework and assignments and in-class examinations (Ross, Niebling, & Heckert, 1999). Students are often penalized due to simple math errors, mistakes in reading the problem, and time constraints during exams even if the students know the material and can correctly apply mathematical formulas. More than 66% of civil engineering students at the University of Pittsburgh at Johnstown reported that time constraints and math errors were the sources of their mistakes on an exam (Murad & Martinazzi, 2003). Student anxiety during a time-constrained examination may also negatively impact the student's performance. Less-anxious students score better on intelligence tests than students with high anxiety (Sarason & Mandler, 1952; Sarason, Mandler, & Craighill, 1952). Stowell and Bennett (2010) argue that test anxiety is comprised of three primary components: affective (physiological arousal, emotional), cognitive (worry), and behavioral (procrastination, avoidance). On the other hand, students may pass a course without demonstrating mastery of the material. Students can often earn partial credit on exams by writing some correct information. If a large number of students perform poorly on an exam or during the semester, many instructors convert the unacceptable grades into acceptable grades (Gordon & Fay, 2010). Curving grades may overstate student performance even if students have not mastered the course content.

The Internet has revolutionized access to information, and engineering colleges are increasingly offering online courses for students (White & Hammer, 2000). Enrollment in online courses has risen exponentially in the United States, and almost 13% more students were enrolled in at least one online course in 2006 than in 2005 (Allen & Seaman, 2008). Students rely on the Internet to submit assignments, find answers for questions, work together in groups, and research course projects. Research on computer-based adaptive learning and the best ways to deliver online courses, engage distance-learning students, and assess student performance has and will continue to proliferate (Vandewaetere, Vandercruysse, & Clarebout, 2012). Student performance in online courses and with online examinations has been better or at par with the performance in traditional classes and exams. Students enrolled in an online course in psychology in Texas Tech University outperformed students in the traditional class environment (Maki, Maki, Patterson, & Whittaker, 2000). The results of unsupervised online quizzes in a medical physiology course demonstrate that the students who chose online quizzes performed better in the semester (Kibble, 2007). The performance of students taking a business course exam in the classroom and in the proctored setting for online exam received similar scores (Alexander, Bartlett, Truell, & Ouwenga, 2001).

Online examinations—regardless of whether or not the class is delivered online—can also facilitate providing multiple attempts for students to pass an exam. Enabling students to learn from their mistakes on an exam may promote learning and mastery of the subject material and reduce test anxiety (Stowell & Bennett, 2010). Students learn from their mistakes if they are allowed multiple attempts to pass an exam. Examinees achieved higher scores on the Graduate Record Examination if they took it a second time (Kingston & Turner, 1984). As the number of attempts of the same version of exam increase, an examinee's scores improve (Wolkowitz, 2011). One explanation for better scores with multiple attempts is that the examinee becomes familiar with the testing format and feels less stressful and anxious while taking another attempt of the test (Terry, 2015). In-class testing also limits the use of computer technology to solve problems, but professionals heavily rely on computers to solve their engineering problems. Traditional in-class examinations may be assessing student performance that bears little reality to how they will do their work in a professional environment.

This article presents and studies the use of online testing modules for a large-enrollment engineering course, engineering economics, during the Fall semester 2017 at a large public university. The online testing modules are designed to address many of the challenges discussed in the preceding paragraphs. The online testing modules allow for multiple attempts so that students do not feel anxious when attempting to pass the online tests and to allow them to learn from their mistakes. The modules encourage the use of Excel and simulation software to solve problems to represent how problems are solved in the professional world. Testing questions are randomized so that it is very unlikely that a student will receive the exact same question even if the student takes an online test dozens of times.

Passing an online testing module requires the student to answer almost every question correct and no partial credit is given. Students are unable to pass a test or pass the course without demonstrating mastery of the material.

This article analyzes how students perform for this type of assessment procedure and the students' opinions about the online testing modules. Specifically, this article focuses on assessing the students' preferences between online testing and in-class examinations, their levels of anxiety, and their engagement in the course with online testing modules. The article quantifies the students' performances for the online testing modules and tests if the student's major or if the course delivery method impacts a student's performance on the online testing modules. Finally, since the online tests allow for multiple attempts, we can test if students are learning through these additional attempts and how quickly that learning occurs.

The uniqueness of this article is the presentation and analysis of a new assessment procedure for a large-enrollment engineering course. To our knowledge, the development and use online testing modules with randomized questions that allow for unlimited attempts has not been written about in engineering education. A survey of students in the course provide insight into students' opinion on this assessment procedure compared to traditional in-class examinations and homework. Statistical analysis of student performance and of the students' improvement in the testing modules enable us to quantify the learning effect of repeating a testing module.

The rest of this article is as follows. Section 2 describes the innovative online testing modules used in the course. Section 3 displays the results of using the online tests and performs statistical analysis on student performance. Section 4 discusses the results and insights gained from conducting this study. Concluding remarks appear in Section 5.

Methodology

The online testing modules were developed and used for the 2017 Fall semester section of the engineering economics. The main learning outcome of this course is for students to correctly apply economics principles to engineering problems. The course covers topics such as the time value of money, net present worth analysis, cost and profit analysis, inflation, and uncertainty and risk. The course is offered in each semester (Fall, Spring, and Summer), and more than 800 students from across the College of Engineering take the course during the academic year. The course is offered as an in-class course and as a distance-learning course.

The Fall 2017 section that experimented with online testing modules included students enrolled in the in-class and the distance-learning sections. Although the class was taught in front of students in the classroom, all the lectures were recorded and were available for viewing to both 87 distance-learning and 155 in-class students. 242 students completed the course. Students with different engineering majors enrolled for this course, and 65 students were industrial engineering (IE) majors.

The grades for this course in the Fall 2017 section depended on passing the online testing modules and responding to two online surveys. No in-class exams, assignments, or final exam were given. Instead, the class evaluated the students through 7 required online testing modules plus 2 additional testing modules. Each of the 7 required testing modules contained 6 to 8 questions. All the questions required applying one or more formulas to solve the problem, and students entered their answer in a single text box. The solutions allowed for +1% of the correct answer in order to avoid penalizing the students for rounding error. Each testing module had a time limit between 90 and 120 minutes so that student could not keep the test open indefinitely.

Questions for each testing module were randomly selected from a large database to ensure that each student had a new question on his on her test. Figure 1 depicts the structure for a testing module with 7 calculation questions (i through vii). Each question randomly chosen from several different problems, labeled as A through D in Figure 2.1. The problems within each question covered the same topic and had the same difficulty. For example, each of the four problems A through D might require students to answer questions about amortized loans. Each problem had 100 different versions where each version had the same text but different numbers.

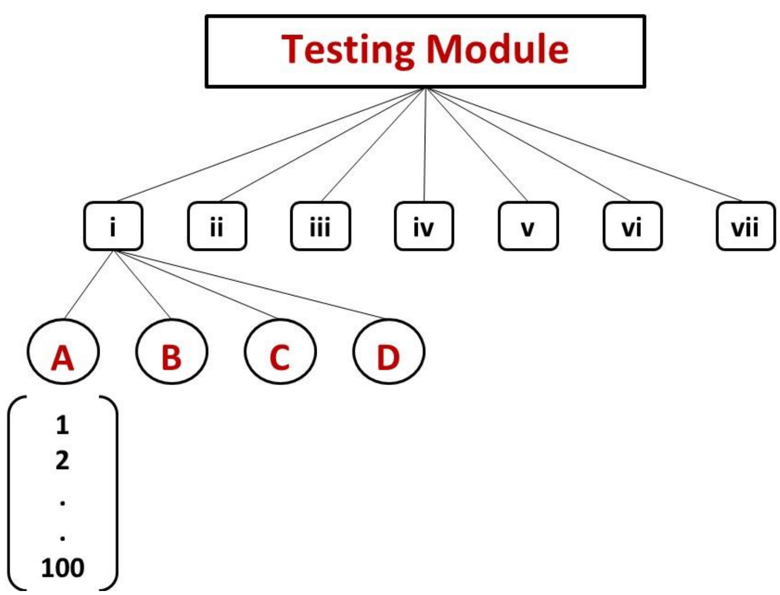

Figure 2.1 Structure of testing modules

For example, Figure 2.2 depicts a question from testing module 4, and this problem appeared as the one of the first questions in testing module 4. The same problem had 100 versions by changing the numerical values depicted in the bold font. Thus, each question had approximately 400 to 600 different possibilities (i.e., 4 to 6 problems with 100 versions each). A student who takes an online testing module several times answered the same problem multiple times, but each time the problem had different numerical values.

Consider the financial data for the following project. The annual revenue and annual expenses occur at the end of years 1 through **6** and the salvage value is recouped at the end of the **6** years.
Initial investment: $**109000**
Annual revenue: $**31000**
Annual expenses: $**8000**
Salvage value: $**12000**
Project life: **6** years
What is the IRR on this project?

Figure 2.2 Example of a problem in a testing module. The numbers in bold are randomly changed on different versions of this problem.

Because it was very unlikely that a student would ever receive the exact same version of a problem on repeated attempts of the testing module, the class allowed for unlimited attempts of the testing modules. Practicing a skill multiple times helps people retain the information in their long-term memory (Willingham, 2004), and learning is enhanced by well-designed repetition (Thalheimer, 2006). Moreover, a study that discussed effective strategies to improve student learning reports that some students benefit from practicing tests (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). When a student took a testing module in the engineering economics course, he or she received the results immediately after submitting the test. For the first three testing modules, the test results showed which questions were answered incorrectly but did not show the correct answers. Based on feedback from the students, we changed the policy for the remaining testing modules by also depicting the correct answer after a student takes a testing module. Providing the student with the correct answers allowed the student to practice the question offline, learn how to correctly solve a problem, and be prepared to successfully answer the question (with different numbers) in future attempts.

The motivation behind assessing students with online testing modules was to encourage students to study the material, understand and apply the concepts in the testing modules, learn from their mistakes, improve their performance, and demonstrate mastery of the subject. Students could use class notes, the textbook, the Internet, and computer software such as Excel as resources to help them answer questions while taking a testing module. Students were encouraged to ask for help from the instructor and the teaching assistants about questions that they had difficulty answering. Students could ask questions in person or via an online discussion forum through the course's learning management system. A student could also ask another student for help in solving the problems as long as neither student was currently taking a testing module. Allowing for repeated attempts of a testing module and requiring that all questions are answered correctly except for one or two questions (depending on the testing module) motivate students to ask for help much more than traditional in-class examinations. Encouraging students to use these resources helps them master the content and learn how to solve realistic, real-world problems rather than focusing their study on memorizing formulas (Chickering & Gamson, 1987).

Table 2.1 depicts the grading scale used for the course. Successfully passing the 7 required testing modules earned a B+ for the semester. Each of the required testing modules focused on material that was discussed in class. To earn an A- or A, a student needed to pass

1 or 2 additional testing modules, respectively. The additional testing modules asked questions about material in the textbook but that was not covered in class. The additional testing modules, testing module eight and nine only had 3 and 4 questions respectively, but otherwise followed the same structure as the required testing modules. Including additional testing modules incentivized students who wanted to earn an A to learn material on their own. The testing modules successively became available during the semester, and students were required to pass the testing modules by the last day of the 16-week semester.

Table 2.1 Grading scale for the semester

Grade	Number of testing modules
A	Pass 7 required testing modules + 2 additional testing modules
A-	Pass 7 required testing modules + 1 additional testing module
B+	Pass 7 out of 7 required testing modules
B-	Pass 6 out of 7 required testing modules
C	Pass 5 out of 7 required testing modules
D	Pass 4 out of 7 required testing modules
F	Pass fewer than 4 out of 7 required testing modules

The methodology used to evaluate the ability of the online testing modules to meet the objectives described previously consists of surveys of the students and an analysis of student performance on each attempt for the testing modules. The course required that students respond to a mid-semester survey and an end-of-semester survey with questions related to the students' opinion about the testing modules, their level of anxiety, and their motivation. The survey also allowed students to offer written feedback about the online testing modules. Responding to these surveys was mandatory and failure to take these surveys reduced a student's grade by one letter mark (e.g., from a B+ to a B). The online testing modules automatically recorded every attempt. This record allows us to evaluate and do statistical analysis of how a student improves by retaking the testing module and having multiple attempts of responding to the same problem.

Results

Results consist of the students' responses to survey questions, the students' performance on the testing modules, and their performance on individual questions. The surveys provide insight into the effectiveness of the testing modules and students' acceptance of this assessment method. The students' performance on the testing modules helped us evaluate if students learn from the testing modules and if the student's major and method of course delivery impacts performance. The students' performance on the testing modules and on individual questions are analyzed to assess if students perform better with multiple attempts at a testing module. A statistical test on student performance on individual questions shows how students improve and learn as they encounter similar questions.

Survey Responses

Students were required to respond to both a mid-semester survey and an end-of-the-semester survey, which allows us to analyze if the students' opinions change during the semester. The two surveys, both asked students questions about their engagement with the course, their learning, and their preferences between online and in-class examinations. Table 2.2 depicts the results of the mid-semester survey taken by 235 students. Questions focused on student anxiety, their engagement with the course, and their motivation to perform in the course.

Table 2.2 Responses of students on the mid-semester survey (percentage of survey respondents)

Question	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Anxious during in-class exams	36.9	43.4	11.2	7.3	1.3
Anxious during online exam	3.4	8.1	15.7	48.9	23.8
Engagement in course with online test	17.9	29.8	31.5	16.6	4.3
Better learning of the material	20.4	28.5	34.9	12.8	3.0
Prefer online testing	40.0	39.2	14.9	4.7	1.3
Less motivation due to no deadlines	6.4	22.2	24.8	32.1	14.5

Table 2.3 shows the results of the survey at the end of the semester on the basis of 233 responses. Some questions were repeated such as whether students preferred the online testing modules, whether they were more engaged with the online testing modules, and whether they learned the material better with the online testing modules. Other questions were added to understand if the students were motivated to pass the additional modules in order to earn an A; if having required deadlines for the testing modules would be a good idea; and if the students would recommend online testing modules for another courses.

Table 2.3 Responses of students on end-of-semester survey (percentage of survey respondents)

Question	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Engagement in course with online test	16.3	23.6	31.3	18.5	9.9
Better learning of the material	17.2	30.9	27.0	27.0	8.2
Prefer online testing	37.3	35.2	14.6	7.3	4.7
Requirement of deadlines to pass in reasonable time frame	23.6	39.9	18.9	13.3	3.0
Deadlines help to perform better	15.9	31.8	27.5	20.2	3.9
Motivated to pass additional modules for an A grade	48.5	18.9	17.2	9.4	4.7
Recommend the assessment method to other courses	21.0	34.3	24.9	11.6	6.4

The surveys indicate that students seemed to prefer the online testing modules although the percentage of students who prefer the online testing modules declined slightly by the end of the semester. According to the mid-semester survey, 79.2% of the students agreed or strongly agreed with the statement that they prefer online testing. That percentage fell slightly 72.5% in the end-of-the semester survey. A little less than half of the students (47.7%) agreed or strongly agreed with the statement that they are more engaged with the course due in the online testing modules in the mid-semester survey than in other courses with traditional assessment. Only 39.9% of the students agreed or strongly agreed with that statement in the end-of-the-semester survey. The mid-semester survey revealed that 48.9% of the students agreed or strongly agreed that they were learning the material better with the online testing modules compared to traditional examinations and homework, and almost the same percentage (48.1%) agreed or strongly agreed in the end-of-the-semester survey.

A key difference between the online testing modules and traditional examinations is the ability of students to retake the online testing modules as many times as they wish. Allowing students to fail, learn from their failures, and then succeed should lesson student anxiety. The mid-semester survey revealed that 80.3% of the students agreed or strongly agreed with the statement that they are anxious during in-class exams. However, only 11.5% of the students agreed or strongly agreed with the statement that they are anxious during the online testing modules.

The survey at the end of semester shows that most students (67.4%) were motivated to pass the two additional testing modules to earn an A for the semester. As the semester progressed, we noticed that many students were procrastinating and leaving all the testing modules for the last few weeks in the semester. The end-of-the-semester survey asked students if having deadlines would be better. Sixty-four percent of the students agreed or strongly agreed that having deadlines for the testing modules would have helped them to pass testing modules in more reasonable time frames by the end of the semester. When asked if students were still motivated to take the testing modules without deadlines, 28.6% of students agreed or strongly agreed that they were less motivated to take the testing modules on their own by the mid of the semester. In general, students at the end of the semester favored online testing modules, and 55.3% of students agreed or strongly agreed that they would recommend this method of assessment for other engineering courses.

Student Performance on Testing Modules

Table 2.4 depicts the number of testing modules that the students passed. More than half of the class reported that they were motivated to learn the material on their own and take the additional testing modules and 52% students earned an ‘A’ or an ‘A-’. Seventy-seven percent of the class passed at least 6 testing modules (earning at least a B- in the course). However, 17 students, or 7% of the class, passed three or fewer testing modules and failed the course. Students who did not take the surveys were given one grade less; however, Table 4 does not consider that grade reduction.

Table 2.4 Distribution of students passing the testing modules

Number of testing modules passed	Letter grade	Number of students	Percentage of students
9	A	119	49
8	A-	7	3
7	B+	6	2
6	B-	55	23
5	C	26	10
4	D	12	5
< 3	F	17	7

Table 2.5 shows the topic of each testing module, the percentage of students in the class who successfully passed each testing module, and the percentage of students who attempted the testing module at least once and passed it. The third column in Table 2.5 shows the percentage of students who passed the testing module divided by the total number of students enrolled for this course, and the fourth column reports the percentage of students who passed the testing module divided by the total number of students who attempted the testing module at least one time. Testing modules 3 and 7 were the most difficult modules, and students had less time to pass testing module 7 because the material was taught at the end of the semester. Testing module 7 focused on decision making with uncertainty and required the students to build Monte Carlo simulations within Excel to answer some of the questions. Seventy-three percent of the class passed this testing module. Only 50% of the class passed the second additional testing module, and a large fraction of the class never attempted this

testing module because this testing module only benefited a student if the student had already passed the seven required testing modules. Over 80% of the students who attempted the additional testing modules passed the additional testing modules.

Table 2.5 Passing rate per module

Testing module	Topics	Percentage of students in the course who passed the testing module	Percentage of students who passed the testing module given the student attempted the testing module at least once
1	Interest rates, economic equivalence	96	96
2	Non-annual compounding, debt repayment	96	97
3	Present-worth analysis, annual-equivalent worth	88	91
4	Rate of return, cost concepts	92	96
5	Project cash flow, depreciation, taxes	98	94
6	Inflation	81	89
7	Uncertainty, simulation, risk	55	74
Additional testing module 1	Economic service life and replacement decisions	57	94
Additional testing module 2	Public sector and cost benefit analysis	50	83

Engineering economics is required course for IE majors at the university, and many students in other engineering majors take this course as a technical elective. The Fall 2017 course was offered as an in-class and via a distance-learning platform. Students who took the course via distance learning watched videos of the class lecture online and could ask questions via the course learning management system and email. The students’ performance can be used to assess if the student’s major and the delivery mode is correlated with the number of testing modules that student passes. Out of the 242 students enrolled for this course, 57 students were in-class students majoring in IE, 98 students were in-class students majoring in a subject other than IE, 8 students were distance-learning students majoring in IE major students, and 79 students were distance-learning students not majoring in IE. Figure 2.3 shows the distribution of students earning the grades from A through F categorized according to major and the course delivery method.

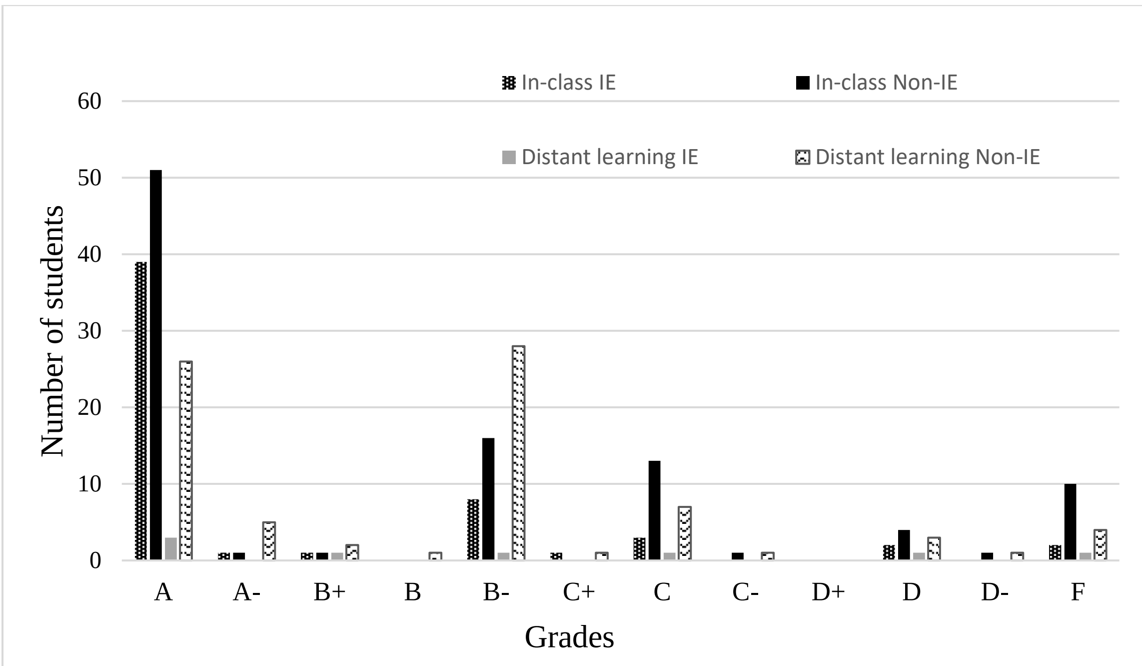

Figure 2.3 Distribution of grades by major and course delivery

The distribution of grades seems to indicate that the student's major and the method of delivery impacts the student's grade. The model to assess the influence of the student's major and the course delivery method is:

$$y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$$

where $y_{ijk} = \{0,1,2, \dots ,9\}$ is the number of testing modules passed, μ is the overall mean for all students, α_j is the main effect of the student's major where $j =$ {IE, non-IE}, β_k is the main effect of the delivery method $k =$ {in class, distance learning}, $(\alpha\beta)_{jk}$ is the interaction,

ϵ_{ijk} is a Gaussian error term, and i represents an individual student. Since each effect has two levels, we arbitrarily assign + for α_{IE} and – for $\alpha_{\text{non-IE}}$ and + for $\beta_{\text{in class}}$ and - $\beta_{\text{distance learning}}$. Table 2.6 shows the average number of testing modules passed by students (i.e., the response variable) divided into categories by major and course delivery. The main effects are estimated as the difference in the average number of modules passed. The effect of the interaction is measured by the difference between the average modules passed when both variables are + or both are – and when the two variables have different signs. The table provides the estimates of the effects of the major of a student and the course delivery method they chose which might have helped them perform better in the course. Although the differences in the average numbers of modules passes are relatively small (less than 1.0), it appears that IE students may out-perform non-IE students and that in-class students may out-perform distance-learning students.

Table 2.6 Estimate of main effects and interaction effect

Category	IE	Non-IE	In-class	Distance Learning	In-class IE & Distance Learning Non-IE	In-class Non-IE & Distance Learning IE
Average number of modules passed	7.57	6.69	7.12	6.60	7.12	6.69
Estimate of the effect	0.87		0.52		0.43	

Data corresponding to Figure 2.3 is used to fit the parameters to the model, which results in a two-way ANOVA test with an unbalanced (unequal sample sizes) design. A type III ANOVA test is appropriate to assess the statistical significance of the parameters because type III tests for the significance of a main effect given the other main effect and the interaction (Herr 1986; Langsrud 2003). Table 2.7 depicts the results of this ANOVA test. The main effect for the delivery method is significant at the 10% level, but neither the main effect for the major nor the interaction term is significant.

Table 2.7 ANOVA test with interaction term

	Sum of squares	Degrees of freedom	F-value	p-value
Student mean μ	312.5	1	56.5	10^{-12}
Major α_j	1.53	1	0.277	0.599
Delivery method β_k	17.0	1	3.08	0.0808
Interaction $(\alpha\beta)_{jk}$	12.4	1	2.23	0.136
Error ϵ_{ijk}	1316.1	238		

Since the interaction term is not significant, an ANOVA test can be conducted under the assumption that there is no interaction effect, i.e., $(\alpha\beta)_{jk} = 0$. Table 2.8 depicts the results of the ANOVA test without an interaction effect. Under this model, the main effect of the major is significant at the 5% level, but the main effect of the delivery method is not significant.

Table 2.8 ANOVA test without interaction effect

	Sum of squares	Degrees of freedom	F-value	p-value
Student mean μ	1759.56	1	316.6	$< 10^{-15}$
Major α_j	23.11	1	4.1578	0.0425
Delivery method β_k	5.37	1	0.9670	0.326
Error ϵ_{ijk}	1328.42	239		

Student Improvement with Multiple Attempts

A major motivation behind evaluating the students with the online testing modules is that a student could take a testing module multiple times and improve his or her performance during those multiple attempts. During the semester, students often opened a testing module and closed the module without answering any questions. Since students encountered no penalty for not answering any questions and they could take a testing module an unlimited number of times, many students got in the habit of searching for instances of a testing module with questions very similar to questions they had seen earlier or with questions they knew how to answer. Recall that a single question in a testing module had between three and seven problems and each problem had 100 different versions. Each of the versions within a problem had the same text but the numbers were changed. Students strategically looked for problems that they had previously solved. To assess if student performance improved on multiple attempts of the same testing module and multiple attempts of the same question, we removed all attempts in which a student did not answer any question in the testing module. We define an attempt in these results as a testing module in which a student entered an answer for at least one question.

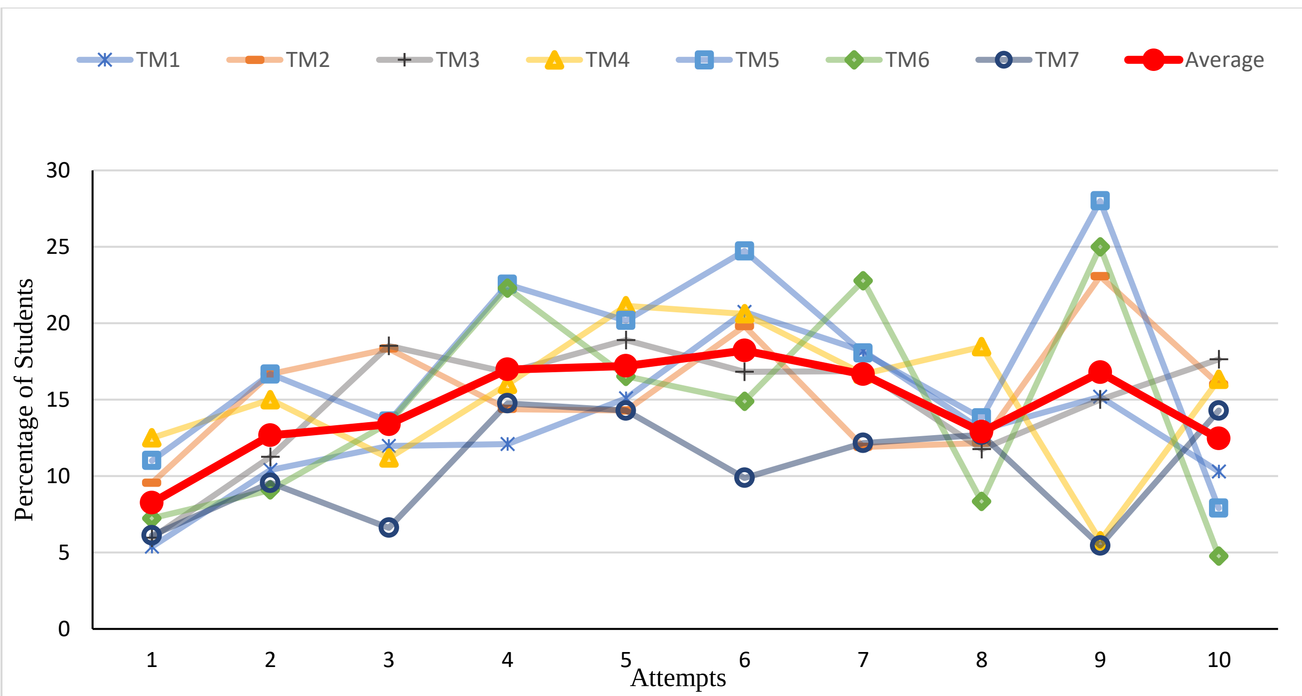

Figure 2.4 Percentage of students who passed a testing module (TM) per attempt

Figure 2.4 shows the fraction of students (as a percentage) who passed a testing module on an attempt divided by the total number of students who attempted the testing module that many times. Since students who pass a testing module do not need to take the testing module again, the total number of students who attempted a testing module decreases with each attempt. The figure shows the success rate for the first ten attempts. However, some students took more than 15 attempts to pass a testing module. The figure shows that approximately 5% of all the students passed testing module 1 on their first attempt. Of the remaining students who attempted testing module 1 a second time, 10% of those students passed testing modules 1 on their second attempt. Of those students who took that testing module a third time, 12% passed it on their third attempt. The bold line with circles depicts the average percentage of students who pass a testing module on the given attempt.

The average passing rate improves from the first attempt to the sixth attempt, and the passing rate for most of the testing modules usually improves during the first six attempts. The improvement during the first four attempts is especially noticeable, but the passing rate is relatively constant from the fourth attempt to the seventh attempt. After the seventh attempt, the passing rates for the modules do not exhibit a trend but are quite variable. The notable variability occurs in part because a large number of students had already passed most testing modules by the seventh attempt so the denominator on which the percentage is based is fairly small.

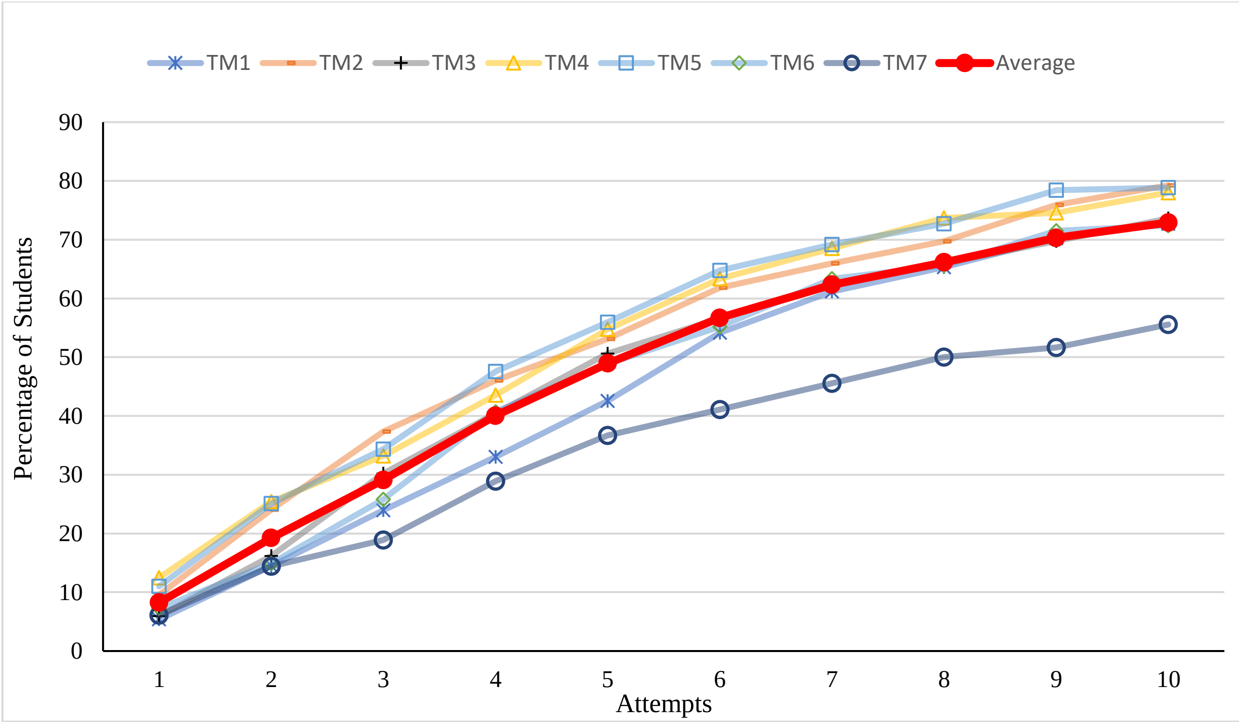

Figure 2.5 Cumulative percentage of students passing a testing module (TM) per attempt

Figure 2.5 shows the cumulative percentage of Figure 4, or the cumulative percentage of students who have passed a testing module by a given attempt. Five percent of students in the course passed testing module 1 on the first attempt, 14% passed the testing module on the first or second attempt, and 24% passed testing module 1 on the first, second, or third attempt. Similarly, for testing module seven, 6% passed on the first attempt, 14% passed on the first or second attempt, and 19% passed on the first, second, or third attempt. The bold line with the circles depicts the average percentage of students who pass a testing module by a given attempt. The graph shows that on average 10% more students pass with each additional attempt through the first five attempts, but the growth rate slows to approximately 5% from the sixth through the eighth attempts. Testing module 7 was the most challenging, and less than half the class had passed the testing module within the first six attempts.

Student Improvement with Multiple Attempts of Same Question

Examining student performance on each individual question can provide further insight into how student's improve by learning how to answer the same question even if the numbers within the question are changed. The logistic regression model is used to estimate the log-odds of answering a question correctly:

$$ln\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

where p is the probability of answering a question correctly, x_1 is the attempt number for a question, x_2 is the total number of attempts by a student for the question, x_3 is a categorical variable that is a unique identifier for each question the testing module, b_0 is the intercept, b_1 is the coefficient for the attempt number, b_2 is the coefficient for the total number of attempts, and b_3 is the coefficient for each unique identifier for a question. The ratio $\frac{p}{1-p}$ is the odds of answering a question correctly. If $b_1 > 0$, the model estimates that the student's probability of answering a question correctly increases with each attempt.

Table 2.9 depicts the results of the logistic regression for each of the seven required testing modules. (Since x_3 is a categorical variable, the coefficient changes for each category and is excluded from the table.) The logistic regression model is significant (p-value < 0.01) for each of the seven testing modules. The parameters b_1 and b_2 are also significant in each testing model, and $b_1 > 0$ and $b_2 < 0$. Both the attempt number and the total number of attempts are significant to estimate the probability of answering a question correctly, but they have opposite effects. The probability of answering the question correctly increases with each attempt. The probability of answering the question correctly decreases as the total number of attempts increases.

Table 2.9 Logistic regression results for student attempt per question

Parameter	Testing module						
	1	2	3	4	5	6	7
χ^2 (Model)	3.43×10^3 **	1.6×10^3 **	2.35×10^3 **	1.75×10^3 **	1.89×10^3 **	2.26×10^3 **	1.15×10^3 **
b_0	1.029**	0.964**	1.519**	1.120**	0.351*	1.277**	0.582**
b_1	0.317**	0.141**	0.135**	0.260**	0.260**	0.215**	0.125**
b_2	-0.227**	-0.295**	-0.179**	-0.256**	-0.195**	-0.103**	-0.106**

Note: * indicates significance at the 10% level, ** indicates significance at the 1% level

Table 2.10 displays the probabilities of answering the question correctly if it appears exactly one, two, or three times. The probability of answering the question correctly decreases as the total number of attempts increase, but the probability of answering the question correctly increases with each attempt. For testing module 1, if a student only attempts a question once, the probability of answering it correctly is 0.75. If a student attempts a question twice, the probability of answering the question correctly on the first attempt is 0.71 and the probability of answering it correctly on the second attempt is 0.77. If a student attempts a question exactly three times, the probability of correctly answering the question increases from 0.66 to 0.73 to 0.79 for attempts one, two, and three.

Table 2.10 Probability of answering a question correctly as a function of attempts

Attempt	Total	Testing module						
number x_1	attempts x_2	1	2	3	4	5	6	7
1	1	0.754	0.692	0.814	0.755	0.603	0.800	0.646
2	1	0.709	0.626	0.785	0.704	0.555	0.783	0.622
2	2	0.770	0.659	0.807	0.756	0.618	0.818	0.651
3	1	0.661	0.555	0.754	0.648	0.507	0.765	0.596
3	2	0.729	0.589	0.778	0.705	0.571	0.802	0.626
3	3	0.786	0.623	0.800	0.756	0.634	0.834	0.655

Table 2.11 shows the odds ratio for the attempt number b_1. The odds ratio shows the percentage increase in the odds of answering the question correctly with each attempt. Questions in testing modules 1, 4, 5, and 6 have the largest odds ratio. Each attempt increases

the odds of answering the question correctly by 24% or more for these testing modules. Questions in testing modules 2, 3, and 7 all have odds ratio between 1.13 and 1.15, which indicates that each attempt increases the odds of answering the question correctly by 13-15%.

Table 2.11 Odds ratio for the attempt number b_1

	Testing module						
	1	2	3	4	5	6	7
Odds ratio	1.37	1.15	1.14	1.30	1.30	1.24	1.13
Percent increase in the odds of answering question correctly for each addition attempt	37.3	15.1	14.4	29.7	29.7	23.9	13.4

Discussion

This new assessment of online testing modules follows three dimensions: (i) students' preference and engagement via the survey, (ii) the performance of the students though the distribution of grades based on the number of testing modules passed, and (iii) student improvement for each testing module and for each question within a testing module. The survey results indicate that students largely prefer the online testing module to traditional homework and in-class examinations. The mid-semester survey (Table 2.2) revealed that 80% of students agreed or strongly agreed that they feel anxious during in-class test as opposed to 12% students who felt anxious while taking online testing modules. These data contrast with the results of a study in a psychology course where students reported same levels of test anxiety and performed comparably well on the single attempt online exams and in-class exams (Stowell & Bennett, 2010). Since a student could take an online testing module as many times as he or she wanted, we would expect the student's anxiousness is a lot less than during an in-class examination, which typically does not allow a student to retake an examination.

Seventy-nine percent of the students preferred the online testing modules according the mid-semester survey, and 73% of the students preferred these testing modules according the end-of-the-semester survey. Students may have preferred the testing modules because they feel less anxious while taking the test. They could take the test anywhere in whatever surroundings (e.g., dorm, library, computer labs) and at any time (day, night, weekday, weekend) that they wished. The ability to take the tests multiple times relieves the stress to pass it on the first attempt. The results from this course are similar to a study about Mallard, an asynchronous Web-based assessment program, in which 97% of the students found that Mallard was beneficial for learning the coursework and they were satisfied with it (Desouza & Fleming, 2003). Summers, Waigandt, & Whittaker (2005) found, however, that students were less content with the online statistics course than the course with a traditional assessment. The difference between their result and the results of this current study could be due to the flexibility with the online testing modules and that we were not explicitly comparing a student's experience between taking a course online versus taking a course in person.

The students' opinions about the online testing modules changed during the course of the semester. The preference for the online testing modules decreased by approximately 6% from the middle of the semester to the end of the semester. In the mid-semester survey, 48% of the students believed the online testing modules helped them remain more engaged with the course. That percentage drops to 40% in the end of the semester. Students may feel more engaged with these type of testing modules because it gives them an opportunity to practice in a non-classroom setting without time constraints (Desouza & Fleming, 2003). One reason that many students may have felt less engaged with the online testing modules and why that engagement and preference for the testing modules decreased during the semester is because the testing modules were challenging. Many students did not realize how many times they would need to take a testing module in order to pass it, and many students indicated their frustration with not passing a testing module because they made a single mistake. This frustration was more evident at the end of the semester than during the middle of the semester.

Having deadlines to pass the testing modules may help with engagement in course and remove some of the frustration. Since the class had no deadlines during the semester to complete the testing modules, many students procrastinated in taking the testing modules and had to finishing several testing modules during the last few weeks of the semester. In the middle of the semester, only 29% of the students believed they were less motivated to pass the testing modules since no deadlines existed. When the students responded to the mid-semester the first three testing modules were available. Testing modules 1 and 2 covered the basics of the course and were comparatively easier than the remaining modules covered later. At the end of the semester, 64% of the students agreed or strongly agreed that requiring deadlines would be beneficial, and 48% of the students thought deadlines could have helped them perform better in the course.

More than half of the class earned an A- or A by passing 8 or 9 testing modules. The distribution of students passing the testing modules (Table 2.4) shows that student performance was distinctly divided into three categories: excellent, mediocre, and poor or failing. Fifty-two percent performed excellent (A- or A), 35% performed mediocrely (C, B-, or B+), and 12% performed poorly (F or D). Figure 2.3 shows that over 70% of the students who attempted a testing module passed each of the seven required testing modules by their tenth attempt, except for testing module 7. The 17 students who failed the course did not attempt all the test testing modules and generally did not take advantage of multiple attempts. Students were encouraged to discuss incorrectly answered questions with other students, the teaching assistants, and the professor in person or via email. The students who procrastinated until the end of the semester had little time to get help or assistance. We speculate that some of these students tried to attempt all the testing modules in the last week and might have gotten frustrated and gave up.

The ANOVA test without the interaction terms reveals that there is some evidence that IE majors performed statistically better than non-IE majors. Since engineering economics is a course aligned with IE students' interests in mathematical modeling and thinking about costs and benefits, it is our experience that IE students generally perform better in this course whether or not online testing modules are used. However, many non-IE students performed excellent and earned an A or A-. Whether or not students took the course in class or distance learning does not have a statistically significant effect on a student's performance. In general, in-class students seemed to perform better than distance-learning students, but that seems to be because a larger percentage of IE students took the course in class compared with non-IE students.

The results examining student performance on multiple attempts indicate that the students improved as they took a testing module repeatedly although students improved at very different rates. On average, 10% of the students pass a testing module with each additional attempt through the first five attempts (Figure 2.4). Many of the students quickly learned their mistakes in the initial attempts and could pass the testing module in a few additional attempts. Five percent of students pass the testing module on average for each attempt from attempt 5 through 7. These students are learning from their mistakes, but they require more time or more attempts to understand the material. The lower percentage seems to indicate that fewer students belong to this category. Finally, after seven attempts, the percentage of students who pass a testing module any given attempt is very variable (Figure 2.3), and on average only 2 or 3% of students in the class pass a testing module in attempt 8, 9, or 10. One of the reasons for this is that by attempt 7, more than 60% of the class has usually already passed the testing module, and the remaining students are either those who are struggling more with the material or who are not putting in the required effort. It is still encouraging to see that a large proportion of the remaining 40% continue to work on the material and successfully pass five or more of the testing modules.

Our results align with the findings of a study at a major Australian university that showed that regular online quizzes have positive effects on students' learning (Angus & Watson, 2009), but a nursing-admission examination showed no improvements with repeated attempts (Wolkowitz, 2011). The difference between our study and the nursing examination is that students in the online testing module environment receive immediate feedback on the questions they answered incorrectly, and they can work to find out the correct method to answer those questions. When a student sees the same question with different numbers, he or she can use his or her prior work to answer the question. The effect of student improvement on individual questions is modeled using logistic regression. The percent increase in the odds of a student answering a question correctly ranges between 13 and 37% for each additional time the student sees the question. This seems to indicate that students are working to understand how to correctly answer questions that they had previously missed.

Conclusion

With the growing enrollment in engineering majors and engineering classes (Yoder, 2012), identifying the best methods to teach large engineering classes and evaluate students in those classes is becoming important. Online testing could be a method to evaluate students in large engineering courses while promoting learning at individual paces. An engineering economics course was selected to implement a new method of assessment via online testing modules. Each online testing module contains thousands of different questions which are randomly chosen, and it is very unlikely that a student will answer the exact same question with the same numbers. Students have the ability to take a testing module as many times as they want. The standards for passing a testing module are set very high. Students could miss at most one or two questions on a testing module and still pass the testing module. Since questions are repeated with different numbers, students could learn from their mistakes and learn the correct methods to solve a question in order to be better prepared for taking the testing module again.

Results from implementing these online testing modules indicate that students in general expressed a preference for the online testing modules versus homework and in-class examinations. Students felt less anxious with these testing modules and recommended them for other engineering courses. Over half of the class—which consisted of many different engineering majors, in-class students, and distance-learning students—earned an A and A- in the course. Three-fourths of the class earned a B- or better. This indicates that the vast majority of the students worked to pass the testing modules and successfully demonstrated mastery of engineering economics material. Less than 15% of the students successfully passed any of the testing modules on their first attempt. Except for testing module 7, over 70% of the students had successfully passed a testing module by the tenth attempt. This improvement demonstrates that students are learning and improving on their previous attempts.

Future research will compare whether students perform better if they are assessed with online testing modules or with traditional examinations. Some researchers express concerns about cheating during online courses (Watson & Sottile, 2010). Future study can be done by implementing preventive measures for electronic-cheating to study the efficiency of online testing. The experiment with the Fall 2017 course indicates that having deadlines by which to pass testing modules may help keep the students motivated throughout the course and lead to better performance. Creating so many different questions with randomized numbers and loading it into a learning management system requires an extensive amount of time by the instructor and teaching assistants.

The specific way in which these online testing modules are created and used in the course represents a unique way to evaluate engineering students. This method can address many of the deficiencies of traditional in-class examinations (e.g., anxious students, lack of learning from mistakes, cramming for an exam, reliance on partial credit) and can encourage students to continue to work in order to succeed. The online testing modules make use of technology and may align more closely with how students want to learn in the 21st century.

References

Adrian, L. M. (2010). Active Learning in Large Classes: Can Small Interventions Produce Greater Results Than Are Statistically Predictable? *the journal of general education, 59*(4), 223-237.

Alexander, M. W., Bartlett, J. E., Truell, A. D., & Ouwenga, K. (2001). Testing in a computer technology course: An investigation of equivalency in performance between online and paper and pencil methods. *Journal of Career and Technical Education, 18*(1).

Allen, I. E., & Seaman, J. (2008). *Staying the course: Online education in the United States, 2008*: ERIC.

Anderson, L. S., & Gilbride, K. A. (2003). Pre-university outreach: Encouraging students to consider engineering careers. *Global J. of Engng. Educ, 7*(1), 87-93.

Angus, S. D., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology, 40*(2), 255-272.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE bulletin, 3*, 7.

Desouza, E., & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education, 14*(2), 121-134.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Gomes, A., & Mendes, A. J. (2007). *Learning to program-difficulties and solutions.* Paper presented at the International Conference on Engineering Education–ICEE.

Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching, 58*(3), 93-98.

Herr, D. G. (1986). On the history of ANOVA in unbalanced factorial designs. *The American Statistican, 40*(4), 265-270.

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education*, *31*(3), 253-260.

Kingston, N., & Turner, N. (1984). Analysis of score change patterns of examinees repeating the Graduate Record Examinations General Test. *ETS Research Report Series, 1984*(1).

Lansgrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, *13*(2), 163-167.

Maki, R. H., Maki, W. S., Patterson, M., & Whittaker, P. D. (2000). Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. *Behavior research methods, instruments, & computers*, *32*(2), 230-239.

Meade, A. W., & Craig, S. B. (2012). Identifying Careless Responses in Survey Data. *Psychological Methods*, *17*(3), 437-455. doi:10.1037/a0028085

Murad, M. M., & Martinazzi, R. (2003). Mixing Exam Formats to Enhance Examination Learning and Test Taking Skills. *age*, *8*, 1.

Ross, S. E., Niebling, B. C., & Heckert, T. M. (1999). Sources of stress among college students. *Social psychology*, *61*(5), 841-846.

Sarason, S. B., & Mandler, G. (1952). Some correlates of test anxiety. *The Journal of Abnormal and Social Psychology*, *47*(4), 810.

Sarason, S. B., Mandler, G., & Craighill, P. G. (1952). The effect of differential instructions on anxiety and learning. *The Journal of Abnormal and Social Psychology*, *47*(2S), 561.

Stowell, J. R., & Bennett, D. (2010). Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research, 42*(2), 161-171.

Summers, J. J., Waigandt, A., & Whittaker, T. A. (2005). A comparison of student achievement and satisfaction in an online versus a traditional face-to-face statistics class. *Innovative Higher Education, 29*(3), 233-250.

Terry, W. S. (2015). *Learning and memory: Basic principles, processes, and procedures*. New York, NY: Psychology Press.

Thalheimer, W. (2006). Spacing learning events over time: What the research says. *Retrieved March, 21*, 2007.

Vandewaetere, M., Vandercruysse, S., & Clarebout, G. (2012). Learners' perceptions and illusions of adaptivity in computer-based learning environments. *Educational Technology Research and Development, 60*(2), 307-324.

Watson, G. R., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses?

White, R. J., & Hammer, C. A. (2000). Quiz-o-Matic: A free Web-based tool for construction of self-scoring on-line quizzes. *Behavior Research Methods, Instruments, & Computers, 32*(2), 250-253.

Willingham, D. T. (2004). Ask the Cognitive Scientist Practice Makes Perfect, But Only If You Practice Beyond the Point of Perfection. *American Educator, 28*(1), 31-33.

Wolkowitz, A. A. (2011). Multiple attempts on a nursing admissions examination: Effects on the total score. *Journal of Nursing Education, 50*(9), 493-501.

Yoder, B. L. (2012). *Engineering by the Numbers.* Paper presented at the American Society for Engineering Education.

CHAPTER 3. COMPARING THE PERFORMANCE OF STUDENTS WITH TRADITIONAL TESTING AND PERFORMANCE OF STUDENTS WITH ONLINE TESTING IN A LARGE-ENROLLMENT ENGINEERING COURSE

Introduction

Education has progressed from the chalkboard to learning through the Internet. Thanks to technological advancement, teaching and learning has reached to a point where hundreds of books can be downloaded in a small device, lectures can be viewed on personal computers or phones thousands of miles away from the instructor, research groups can discuss material and projects from different parts of the world, assignments can be submitted via emails or web-portals, and students across the globe can take exams. The Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland has introduced selected pilot courses in engineering by enabling students to perform an experiment from anywhere and anytime (Gillet, Ngoc, & Rekik, 2005).

The Internet is emerging as a teaching and learning tool rather than simply facilitating distance education. The Internet can help students learn material on their own and give students confidence in their ability to act as independent learners (Kian-Sam, Abang Ahmad, & Ming-Koon, 2003). Online methods have been adopted as early as in 2002 to conduct many online laboratory experiments in science and engineering (Ammari & Slama, 2006; Salzmann, Gillet, & Huguenin, 2000). Engineering education has been assisted by the Internet by providing e-Journals, documents, and references which can be shared and stored in large numbers for references. Engineering courses are offered online and its enrollment has risen exponentially in the United States (Allen & Seaman, 2008; White & Hammer, 2000).

As technology and the Internet have created new methods for students to learn and for instructors to teach and assess student learning, this phenomenon has raised new questions on the best methods to help students learn course material. The use of well-designed online modules can help students better understand course material (Henson, Fridley, Pollock, & Brahler, 2002). Instructors put a lot of thought and time in developing coursework and lessons that will help students in their future careers, but students often do not retain the lecture material (Lyle & Crawford, 2011). To bolster student learning, it is essential to recognize a student's shortcoming in understanding course material, provide constructive feedback to students, allow for students practice the material, and assess student learning. Designing the curriculum based on the learners' characteristics and modifying the existing instructional design can help students learn more effectively (Passerini & Granger, 2000; Zacharis, 2010). Traditional in-class exams may not be the best way to assess students or to help students learn, especially in the Internet age. For more than a century, research has investigated the use of tests to help students learn rather than just to assess students (Gates, 1917; Jones, 1923; Lyle & Crawford, 2011).

Utilizing computers and other technology in teaching and learning domains can be effective. Students in a psychology class performed significantly better using online quizzes than student who took traditional paper-and-pencil quizzes (Desouza & Fleming, 2003). A study on the first year module in geographical data analysis introduced with computer-based assessment showed that the module evaluation improved and students were content about it (Charman & Elmes, 1998). The former also reported that they were "very satisfied" with online quizzes. Online tests enable students to determine when and where to take their exam, which provides flexibility for the students. Daily online testing conducted in two large introductory psychology classes demonstrated that student performance improved more than the performance of students in a traditional class taught by the same instructor (Pennebaker, Gosling, & Ferrell, 2013).

Practicing is an effective method of learning. Studies done by Butler & Roediger III (2007) to understand the benefits of testing on retention in a simulated classroom setting demonstrated that recall tests after almost one month of teaching and short answer tests can help students retain more material after long intervals. Students received better scores on the Graduate Record Examination when they took it the second time (Kingston & Turner, 1984). Students can receive immediate feedback with online exams, which is usually not possible with traditional exams. Online tests can be used for practicing and providing quick feedback to the students so that they learn from their mistakes. When students practice multiple times with the same or similar versions of a test, their scores improve. Possible reasons could be familiarity with the testing format due to repetition (Terry, 2015; Wolkowitz, 2011).

Anxiety can be a catalyst to poor academic performance. Anxiety among students can include panicking and going blank before test, lack of interest in the course, and feeling helpless while studying it, sweaty palms, and a fast pace of breathing and heartbeats. Engineering students may be particularly susceptible to anxiety (Ruffins, 2007; Vitasari, Wahab, Othman, Herawan, & Sinnadurai, 2010). Traditional in-class exams can increase student anxiety because most engineering courses only have 3-4 exams during the semester, much of the student's grade depends on performing well on each exam. Eliminating worry can help treat test anxiety (Tryon, 1980). Students reported feeling less anxious about taking online tests than in-class tests (Stowell & Bennett, 2010). Moreover, self-regulation helps students to assess their own work and feedback building on the self-studying ability (Nicol &

Macfarlane-Dick, 2006), which could help decrease their anxiety. One challenge with online exams is it makes it easier for students to cheat. (Kennedy, Nowak, Raghuraman, Thomas, & Davis, 2000) report that with the increase of web-based learning and online exams, academic dishonesty will increase.

This article analyzes how the assessment procedure impacts the performance of students in a large-enrollment engineering economics course at a large public university. The article compares students who were assessed via online testing modules with students who were primarily assessed with traditional homework and in-class examinations. The online testing modules contained randomized questions. A student could take a testing module multiple times and never encounter the exact same problem. Students who were assessed with these online testing modules during the 2017 fall semester could take a testing module as many times and they wished until they passed. Students who took the same course during the 2017 spring semester were assessed via weekly homework assignments, a group project, and three in-class exams. The final exam for the spring course was structured as an online testing module as a precursor to the fall semester. This article compares the difference in student performance between students in fall who were assessed via online testing modules and students in spring based solely on their homework, project, and in-class exams. This article also compares student performance in the spring on the traditional exams and on the online final exam.

To the authors' knowledge, little research has examined the effectiveness of using online exams with multiple attempts for learning in engineering classes, and even less research has compared the performance of engineering students with online exams and with traditional in-class exams. The article compares how the same students performed on traditional assignments and exams versus an online final exam. The grades of students who were assessed via online testing modules in fall are statistically compared with the grades of students in the spring. Since some of the same questions were used in the in-class exams in the spring and in the online testing modules in the fall, a unique element of this article compares how students performed on specific questions based on whether the question was asked in an online or in-class test. The fall students may have attempted the same question multiple times, and this article compares the ability of the students to correctly answer a question that they see multiple times with students who see the question one time on an in-class examination.

This article provides important empirical evidence to help answer if students perform better with online tests that can be repeated multiple times versus traditional in-class exams and homework. Section 2 describes the online testing modules for fall 2017 and the more traditional methods of assessing students in the spring 2017 semester. Section 3 presents and analyzes the data on student performance, to include comparing the grades between the two classes and comparing how students answered specific questions. Section 4 discusses the results.

Methodology

The engineering economics courses at Iowa State University typically enroll about 1000 engineering students a year and have multiple sections each semester. Some sections are online classes only, and some sections are in-class sections only. The main learning outcome of this course is for students to correctly apply economic principles to engineering problems. We compare two sets of students with different assessment method to evaluate their performance. Students in fall 2017 had their 100% of grading based on the online testing modules. Students in spring 2017 were evaluated based on their scores on their homework, project, three in-class tests, and an online final exam. This research involved the study of gauging the efficiency of new assessment method in a high-enrollment engineering-course.

Fall 2017

The engineering economics course in fall 2017 had 242 students, which included both in-class and distant-learning students. The class was taught in the classroom and the recordings of each lecture were made available to the distance-learning students. The students' performance was evaluated based on the results of the online testing modules which could be taken multiple times. Since the online testing modules allowed students to use Excel, many of the classroom lectures focused on teaching students how to use Excel and an Excel-based simulation software to solve engineering economics problems.

The grades for this course depended on passing the seven online testing modules with two additional testing modules. Each testing module contained a certain number of questions (usually 7 or 8). All the questions required application of engineering economics formulas to solve, and students entered their numerical answer to each question. Passing a testing module usually required the students to correctly answer all but one question. The instructions for each testing module specified the number of correctly answered questions required to pass. The solutions allowed an answer to be within $\pm$1% of the correct answer in order account for rounding error. Partial credit was not given for any question. Each testing module also required the student to state on his or her honor that he or she had not cheated while taking the testing module. Although students were allowed to use their own computers and could refer to notes and the Internet during these testing modules, they were not allowed to discuss questions with other students while they were taking a testing module. Each testing module had a time limit between 90-120 minutes to prevent students from keeping a testing module open indefinitely.

Questions for the testing modules were randomly selected for each student so that it was very unlikely that a student ever received the exact same version of a question during his or her multiple attempts. For each attempt on a testing module, a question randomly chose among 3-6 different problems. The problems usually covered the same topic and were roughly equivalent in difficulty. For example, the problems might all be questions related to calculating annuities. Each problem had 100 different versions. All versions of the same problem had identical text, but each version had different or randomly selected numbers. For example, if a problem has a student to calculate the present value of an annuity, one version might have an annuity of $2,000 each year for 25 years with an interest rate of 6%, and another version might have an annuity of $3,400 each year for 28 years with an interest rate of 4%. Thus, each question on a testing module had 300-600 unique versions and answers. Having so many different versions ensured that a student could not simply memorize and regurgitate an answer and made cheating more difficult.

Since the standard for passing a testing module was relatively high and students could take each testing module as many times as they needed to in order pass it, this format helped students practice to retain information in their long-term memory (Willingham, 2004). Testing students with exams appears to be a more efficient learning method than other teaching techniques (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Students received the results from a testing module immediately and could practice the problems the missed offline. Well-designed repetition is also an effective learning practice (Thalheimer, 2006). The first four testing modules only told the students which questions they answered incorrectly but did not provide students with the correct answers. Based student feedback, the instruction team changed the testing modules beginning with the fifth module so that students would also receive the correct answers after taking a testing module. Students could re-solve the problems and check to see if their new solutions were correct.

Students were allowed to use textbooks, class notes, the Internet, and Excel while taking a testing a module. They were not allowed to talk with other students or receive help from any individual while taking a testing module. After a student took a testing module, he or she could discuss the questions with other students and receive help from the professor or teaching assistant. Referring to textbooks, practicing the problems with peers, and receiving help from the instruction team allowed students to learn through a method that suited them the best. These additional sources for help combined with repetition help students to master content (Chickering & Gamson, 1987). The motivation behind this type of testing modules was to allow students to study the material, understand and apply the concepts to the questions, learn from their mistakes, retake the testing modules, and improve their performance. Students were highly encouraged to interact with the teaching assistants and the professor for help on questions.

Grades were assigned according to the number of modules passed during the semester (Table 3.1). The seven required testing modules tested material covered in the classroom lectures, and a testing module usually covered about 2-3 weeks of class lecture. Students who wanted to earn an A in the course were required to successfully pass two bonus testing modules. The bonus testing modules asked questions on material that was in the textbook but that was not covered in classroom lectures. Although students could ask the instruction team about questions in the bonus testing modules, the goal of the bonus testing modules was for students to learn material independently of classroom lectures. The bonus modules did not help a student's grades unless he or she successfully passed the seven required testing modules.

Table 3.1 Grading scale in fall 2017

Grade	Number of testing modules
A	Pass 7 required testing modules + 2 additional testing modules
A-	Pass 7 required testing modules + 1 additional testing module
B+	Pass 7 out of 7 required testing modules
B-	Pass 6 out of 7 required testing modules
C	Pass 5 out of 7 required testing modules
D	Pass 4 out of 7 required testing modules
F	Pass fewer than 4 out of 7 required testing modules

Students in the fall semester were surveyed twice—in the middle of the semester and at the end of the semester—in order to understand their opinion about this new method of assessment in an engineering course. Incentivizing students to respond to course surveys and teaching evaluations through the use of grades increases the response rate (Dommeyer*, Baum, Hanna, & Chapman, 2004). In order to encourage a high response rate, a student's grade was deducted by a grade (e.g., from a B+ to a B) if he or she did not respond to both surveys.

Spring 2017

The instructor also taught engineering economics in spring 2017. This section had 162 students, and they were exclusively in-class students. This semester largely had traditional assessments for students: three 50 minutes exams, a final exam, a group project, and eight homework assignments with 5-7 problems each. The second exam had two parts: an in-class part and a take-home part due 48 hours later. Students generally performed poorly on the third exam, so scores were curved on the third exam. The grading scale used for the spring 2017 course is shown in Table 3.2

.

Table 3.2 Grading scale of spring semester

Grade	Percentage range	Grade	Percentage range
A	92.5 - 100	C	72.5 - 76.49
A-	89.5 - 92.49	C-	69.5 - 72.49
B+	86.5 - 89.45	D+	66.5 - 69.49
B	82.5 - 86.49	D	62.5 - 66.49
B-	79.5 - 82.49	D-	59.5 - 62.49
C+	76.5 - 79.49	F	59.49 or below

The final exam in this course was an online exam available for 10 days. The instructor used the final exam to evaluate if testing modules could be used in the following (fall 2017) semester. The online final exam was very similar to the online testing modules used in the fall 2017 class. The final exam had seven questions, and students could take the final exam as many times as they wanted in order to improve their grades. Each question had 4-6 problems, and each problem had 100 different versions. The best score for a student on the final exam was recorded as the final exam score. Table 3.3 depicts the grading scale for the final exam.

Table 3.3 Grading Scale for the final exam in spring

Correct answers	Score	Correct answer	Score
7	100%	3	65%
6	95%	2	55%
5	85%	1	45%
4	75%	0	0%

Our hypothesis is that using the online testing modules with repeated attempts to assess student performance in a large-enrollment engineering course helps students learn the material than compared with traditional in-class exams. In order to assess the validity of this hypothesis, we compare: (i) the performance of students in spring 2017 on traditional in-class exams and the online final exam; (ii) the performance of students in spring 2017 with the traditional exams and homework assignments and the performance of students in fall 2017 with the online testing modules; and (iii) the ability of students in spring 2017 and to fall 2017 to answer the same question.

Results

Results consist of comparing the performance of students in the spring semester who were assessed traditionally and had an online final exam, comparing students assessed traditionally in spring 2017 and student's performance on testing modules in fall 2017, and the responses of students who took the testing modules. The student's performance in spring on the in-class exams and the online final exam enables us to compare the same student's performance on two different assessment procedures. Comparing students in the spring with students in the fall is a between-group design, and we compare the overall grades and performance on individual questions. The fall 2017 students' responses to the survey questions provide a way to assess if students believe they learn better with the online testing modules.

Spring 2017 traditional exams and Spring 2017 online final exam

Prior to the final exam, students in the spring 2017 course knew their percentage grade based on eight homework assignments, one group project, and three in-class exams. Students could calculate their letter grade based on their percentage using Table 3.2.

. We compare those the percentage grades from the traditional assessments to the percentage grade of students on the online final exam (as shown in Table 3.3). The students could calculate exactly how their percentage grade on the final exam would be combined with their prior grade to translate to their semester letter grade.

Figure 3.1 displays the performance of students from their traditional assessment and final online exam. The data points above the dotted line show the number of students who performed better on the final exam, and the data points below the dotted line show the number of students who performed worse on the final exam. The mean difference between the traditional assessment and the online final exam was -0.167, but a matched-pairs t-test finds that the mean difference is not significantly different from 0. Many students who had an A or A- before the final exam only needed an 85% (or in some cases a 75%) on the final exam to maintain their A or A- for the semester grade, so they were not incentivized to continue to take the final exam to earn a 95% or 100%. Many students who were earning a D or C before the final did take advantage or repeating the final multiple times in order to improve their semester grade to a C or B, respectively. Before the final exam, 13 students were earning a D+, D, or D-, and 5 of those students did well enough of the final to improve their grade to a C- or better.

Spring 2017 and fall 2017 letter grades

To compare the performance of students assessed with in-class exams and with online exams over the entire semester, we compare the pre-final grades of 162 students in spring 2017 to the semester grades of 242 students in fall 2017. The spring and fall courses covered the same material in course with the same instructor. We choose the pre-final grades for students in spring 2017 because the final for spring was an online examination.

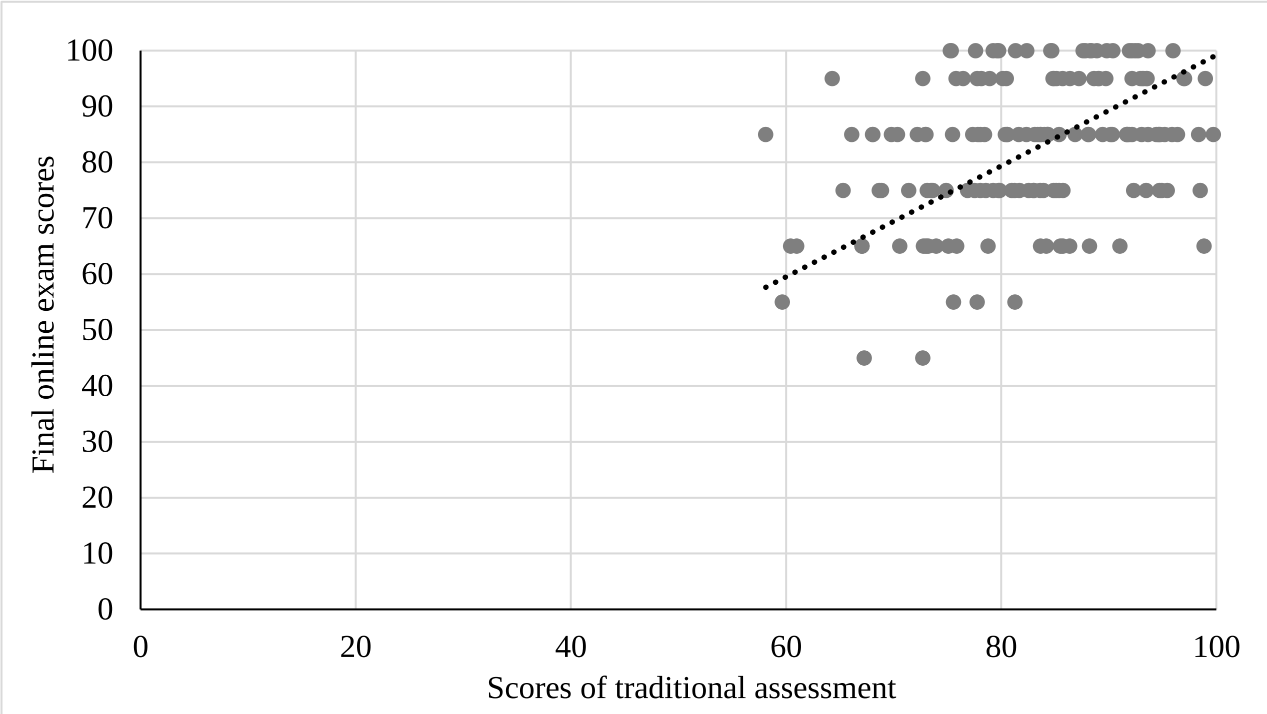

Figure 3.1 Comparison of traditional assessment and final online test evaluation in spring semester

As shown in Figure 3.2, the percentage of students in the fall semester who earned an A is more than twice the percentage of students in the spring semester. The largest percentages for the fall semester grades occur at A, B-, C, and F. The spring semester grades are more evenly distributed across the 12 grades. Students in fall who earned a B, C+, C-, and a D- are those students who did not take one of the two surveys, which negatively impacted their semester grades. The average grade in the fall was a B, and the average grade in the spring was a B-. Our results echo a study at Texas Tech University where students in an online psychology course outperformed students in the traditional test environment (Maki, Maki, Patterson, & Whittaker, 2000). Students in fall semester had the flexibility in taking their tests within the span of the entire semester.

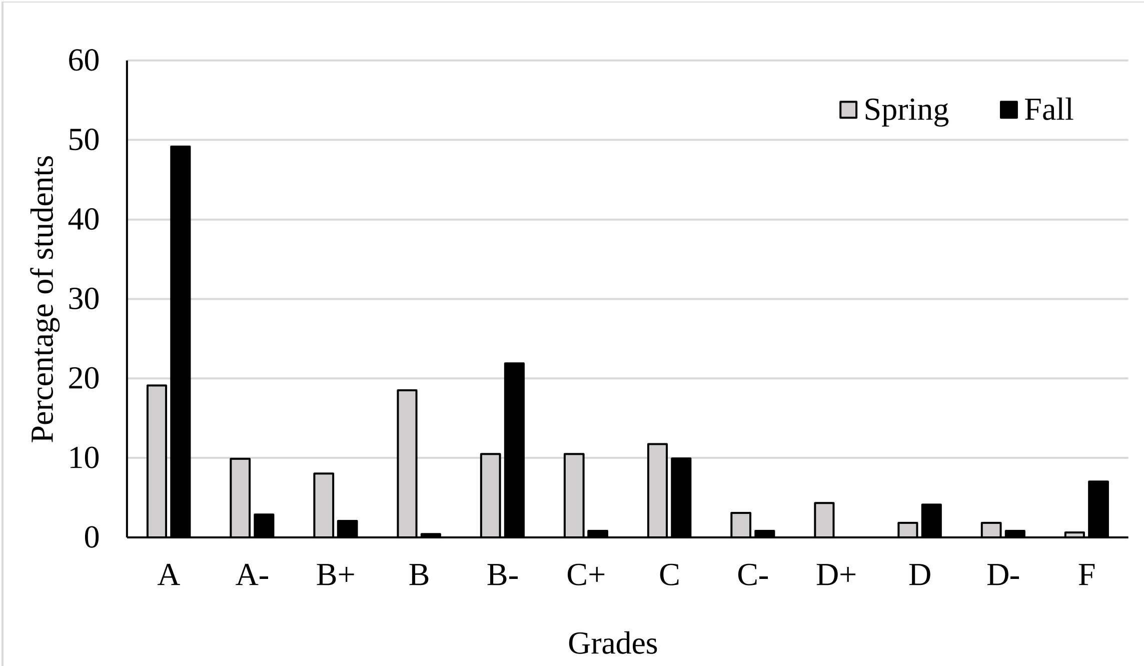

Figure 3.2 Distribution of grades for spring '17 (traditional assessment) and fall '17 (online assessment)

We test to see if the distributions of grades in the fall and spring are significantly different from each other. We test if the grades divided into the 12 categories (e.g., A, A-, B+, B) as depicted in Figure 7 are significantly different and if the grades divided into 5 categories (i.e., A, B, C, D, and F) are significantly different. We conduct a Pearson chi-squared test to evaluate if the difference in grades between the two semesters occurred by chance. Since the expected value in some of the categories is very small (less than 5) we also conduct a Fisher's exact test. Both the Pearson's chi-squared test and Fisher's exact test returned very small p-values (less than 0.01) for the grades divided into 12 categories and for the grades divided into 5 categories. Thus, we conclude that the distributions of grades between the spring and fall semesters are significantly different from each other.

Spring 2017 and fall 2017 question comparison

Eleven problems from the three in-class exams in the spring also appeared on the online testing modules in the fall. These eleven problems were distributed among the seven required testing modules. Students in the fall had to enter a numerical answer that was within 1% of the correct answer in order to have the question correct on the testing module. We compare the percentage of students who solve these problems correctly with the percentage of students in the spring who solved the same problems on the in-class exams.

As discussed earlier, even though the problem in the fall online testing module has the same words, the numbers are varied each time since there are 100 versions of each problem. Since students in the fall could take a testing module as many times as they needed to until they passed and suffered no penalty for failing to pass a testing module, many students opened a testing module to look at the questions and submitted the module without answering any question. They got a score of 0 on that particular attempt, but they could work on the problems offline. They would attempt the testing module again. Thus, when we count the number of attempts on a problem or testing module, we only use “valid” attempts. We define a valid attempt as an attempt on a testing module in which a student entered a numerical answer for at least one of the questions on a testing module. If a student did not enter a number for any question, this attempt was excluded from the count in this comparison.

Figure 3.3 displays the proportion of correct answers for students in the spring and fall. The four column bars in the graph shows proportion of students in the fall who solved the problem correctly on attempts 1-4 (i.e., the first, second, third, and fourth time the student sees the problem on the testing module). The bold line with squares (best attempt) represents the proportion of students who solved the problem correctly on the attempt in which they received their best score on that testing module. This was often the attempt during which the student passed the testing module. Due to the randomness in the online testing modules, a student may not have received that problem on his on her best attempt and he or she may have received a different problem. Those students are not included in the data for best attempt. The grey data line with circles depicts the proportion of students in the spring who solve the problem correctly during the in-class exam.

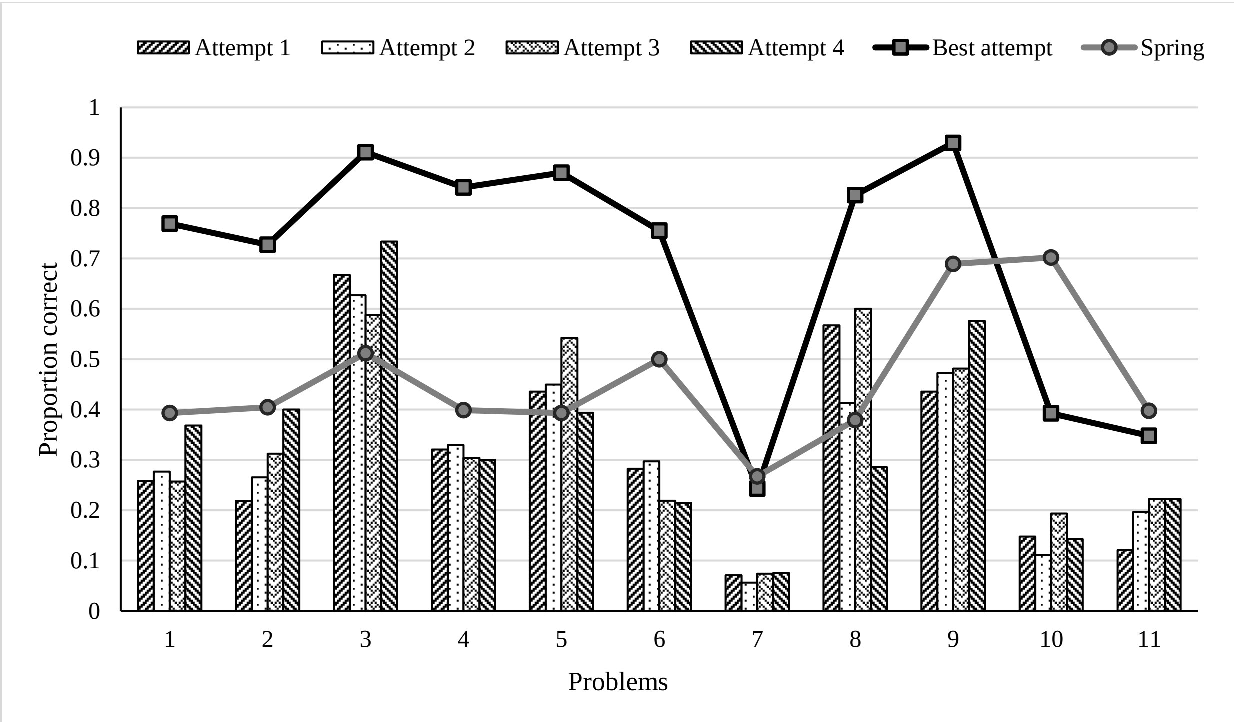

Figure 3.3 Success rate on common problems in fall and spring. Attempts 1-4 and Final attempt correspond to the fall online testing modules.

Figure 3.3 shows that a higher proportion of students in the spring correctly solved the problems than students during their first four attempts in the fall for all but three problems (problems 3,5, and 8). However, the bold line with squares shows the success rate for the best attempt on the testing modules is much higher than in the in-class exam. The column bars for first four attempts increase for many problems on each attempt. This indicates that students are improving as they take a testing module multiple time. The proportion of correct answers to the number of questions in the online testing module on the first attempt is 0.30 and on in-class exam is 0.46; however, the proportion of correct answers on the best attempt for the online testing module is 0.67.

Table 3.4 shows the difference in the proportions of students who answered questions correctly between the spring in-class exams and the different attempts in the fall testing modules. Several students enrolled in the spring semester dropped the course by the middle of the spring, and 168 students answered the first six problem in the spring and 161 students answered the last five problems in the spring. Out of the eleven problems, students did better on the online tests on problem numbers 3, 5, and 8 starting from their first attempt than the in-class exams. The table also depicts if these differences are statistically significantly different from 0. A two-tailed test for population proportions was performed, where the null hypothesis is that the difference between the passing rates equals to 0 and the alternative hypothesis is that the difference is not equal to 0. For the first attempt of the online testing module, the results of only two problems were not significant, for attempt numbers two, three, and four, four, five, and seven problems were non-significant respectively. For the best trial the students had on their online test, only two results were non-significant.

Table 3.4 Difference in success rate between the spring and fall on the same problem

Pro-blem	Number of students in spring course	Proportion of students in spring who solved problem correctly	Proportion in fall who solved problem correctly for a given attempt minus proportion in spring who solved problem correctly *(Number in italics represents the number of students in fall course who attempted question)*				
			1st attempt	2nd attempt	3rd attempt	4th attempt	Best attempt
1	168	0.393	-0.135 ** *151*	-0.116 * *65*	-0.136 (ns) *35*	-0.024 (ns) *16*	0.376 *** *39*
2	168	0.405	-0.187 *** *110*	-0.139 * *49*	-0.092 (ns) *16*	-0.005 (ns) *5*	0.323 *** *22*
3	168	0.512	0.155 *** 132	0.115 (ns) 59	0.076 (ns) 34	0.221 * 15	0.399 *** 45
4	168	0.399	-0.079 (ns) 153	-0.070 (ns) 82	-0.094 (ns) 46	-0.099 (ns) 20	0.469 *** 44
5	168	0.393	0.043 (ns) 179	0.057 (ns) 109	0.150 ** 59	0.001 (ns) 33	0.477 *** 77
6	168	0.500	-0.217 *** 138	-0.203 *** 64	-0.281 *** 32	-0.286 ** 14	0.256 *** 45
7	161	0.267	-0.196 *** 184	-0.211 *** 124	-0.193 *** 81	-0.192 *** 53	-0.023 (ns) 74
8	161	0.379	0.188 *** 141	0.034 (ns) 75	0.221 ** 35	-0.093 (ns) 21	0.447 *** 46
9	161	0.689	-0.254 *** 163	-0.217 *** 91	-0.208 *** 54	-0.114 (ns) 33	0.240 *** 71
10	161	0.702	-0.554 *** 169	-0.591 *** 99	-0.508 *** 62	-0.559 *** 42	-0.309 *** 79
11	161	0.398	-0.277 *** 240	-0.201 *** 61	-0.175 (ns) 18	-0.175 (ns) 9	-0.050 (ns) 46

Note: (ns): non-significant, *: <0.1, **: <0.05, ***: <0.01

Fall 2017 survey response

The instruction team surveyed students in the fall about the use of online testing modules. Students were required to respond to two surveys, but the students' responses were anonymous. Anonymous online surveys may give better responses, both in quality and quantity, than in-class surveys, because the respondents are not worried about identity (Dommeyer et al., 2004). Students responded positively about the new method of testing. Around 79% students preferred online testing by the mid of fall semester (Table 2.2). This might be because students were free to take these tests anytime and anywhere on the campus off campus. Due to taking the tests in a comfortable environment, students felt less anxious while taking the tests. Fifty-five percent of the respondents agreed or strongly agreed with the statement that they would recommend online testing modules as an assessment procedure for other engineering courses (Table 3.5). One-fourth of the survey respondents neither agreed nor disagreed with that statement.

Table 3.5 Responses to the survey questions

I recommend this type of assessment process (online testing modules) for other engineering classes.	Percentage of students
Strongly Agree	21.0
Agree	34.3
Neither Agree nor Disagree	24.9
Disagree	11.6
Strongly Disagree	6.4
No answer	1.7

Twenty percent of students believed this engineering economics course was much better than other engineering courses, and 65% students reported that this course was a little better or at par with other engineering courses (Table 3.6). Overall, the students seemed to be satisfied with the testing modules and the flexibility of online testing modules offered in this course.

Table 3.6 Responses to the survey questions

How good or bad was the course compared with other engineering courses you have taken at Iowa State?	Percentage of students
This course was much better than other engineering courses.	20.2
This course was a little better than other engineering courses	32.6
This course was about the same as other engineering courses	32.6
This course was a little worse than other engineering courses	9.4
This course was much worse than other engineering courses	3.4
No answer	1.7

Discussion

The goal of spring 2017 assessment was to assess the students based on their scores in their homework, exams, and the project. On the other hand, the fall 2017 students were allowed to take multiple attempts of tests so that they could improve by learning from their mistakes. The repeated online testing modules focuses on improving through repetition, which is different than the assessment process in the traditional homework and in-class examinations. Since the assessment methods have slightly different goals, making a straightforward comparison between the two classes is challenging. The results of this analysis and comparing between online testing and in-class exams suggest that the online testing modules may help students learn material better and perform better than traditional homework assignments and in-class exams. The results are not uniform across all the students, however. Comparing the performance of students in the spring on their homework and in-class exams with their performance on the online final exam shows a wide disparity. Some students performed better on the traditional work, and other students performed better on the online exam. Since the best students did not need an A on the final exam to earn an A for the semester, these students were not incentivized to perform their very best on the final. This skews the results. Many students who had performed mediocrely on the traditional assignments took advantage of the online final with multiple attempts to improve their grade. These results are similar to the results of Kibble (2007), which showed that students who chose online quizzes in a medical physiology class performed better in the semester. However, we speculate that students were being strategic while taking the online final exams in spring semester because they knew their in-class assessment grades beforehand.

Improvement in the exam performance because of introduction of Computer-based Assessment suggested that the learning effectiveness of students was also improved (Charman & Elmes, 1998). We compare the grades of students in fall semester who were assessed via the online testing modules to the grades of students in the spring before their final online exam. Prior results show that if the student assessment interface is well created, students can do better with online tests (Ricketts & Wilks, 2002). More than 50% of students in the fall earned an A or A-, compared with almost 30% of students in the spring. Many students who might normally earn a B+ or B through traditional assessment procedures seemed to take advantage of being able to retake the online testing modules in order to earn an A. Many more students in the fall earned an A than those who earned an A- or B+. Since students could improve their grade from a B+ to an A- by passing the two additional modules on material not covered in class, this suggests that students were motivated to learn material on their own to attempt the additional testing modules. Students in the fall had flexibility because they did not have time constraints or deadlines on the testing modules to submit their tests. The availability of practicing the tests multiple times might have comforted the students and reduced the anxiety by regularity in studying (Leeming, 2002; Stowell & Bennett, 2010).

The possibility of retaking a testing module seems to have reduced student anxiety and helped them perform better (Sarason & Mandler, 1952; Sarason, Mandler, & Craighill, 1952). The percentage of students who earned a B- or C+ in the spring was approximately equal to the percentage of students who earned a B- or C+ in the fall. The percentage of students who earned a C in each semester was also roughly equivalent.

Seven percent of students in the fall failed the course compared with less than 1% of the class who failed the class in the spring. More students failed the course with the online testing modules because they procrastinated too much and failed to take advantage of the multiple attempts. Often, engineering students who barely pass engineering courses are able to take advantage of receiving partial credit on examinations. Since the online testing modules offered no partial credit and the standard for passing a testing module was relatively high, these types of students were unable to demonstrate sufficient mastery of subjects to pass the course.

Eleven problems were asked of students in the spring and in the fall. The average percent of problems answered correctly was 46% in the spring and was only 30% on the first attempt in the fall. Since the fraction of students who answered these questions correctly on the online testing modules in their first few attempts were much smaller than the fraction of students in the spring, this suggests that students in the fall did not spend much time studying the material before attempting the online testing modules. Students in the spring studied for the in-class exams and were generally much better prepared to successfully solve these problems correctly. The spring course also provided practice exams with similar types of problems so the students in the spring could practice solving similar problems before encountering these eleven problems on the exam.

Taking more attempts and seeing the same problem (with different numbers) does not substantially improve student performance in the fall, at least during the first few attempts. A higher percentage of students in the spring solved nine out of eleven problems correctly than students in the fall on their fourth attempt. The difference between these percentages on these nine problems was statistically significant at the 10% level on four of these nine problems. One reason why there is not more improvement by students in the fall is that the number of students who attempt one of these eleven problems three or four times is fairly small, especially compared to the total number of students in the class. The best students might only attempt that problem one or two times because they might have been able to pass the testing module fairly quickly.

Students in the fall who solved one of these eleven problems during their best attempt on a testing module had a higher success rate than students in the spring on eight out of the 11 problems. The proportion of correct answers is frequently much higher in the students' best attempt in the fall than the proportion of correct answers in the spring. For example, 40% of the students in the spring solved problem 4 correctly, and 87% of the students in the fall solved problem 4 correctly during their best attempt in the testing module. Of the eight problems for which fall students in their best attempt outperformed students in the spring, the difference in the success rate is statistically significant at the 1% level for all eight problems. Students in the fall improved their performance, and a higher percentage of students in the fall ultimately demonstrated that they could solve engineering economics problems than the percentage of students in the spring. Since spring students only had one opportunity to demonstrate that they could solve these problems, the students in the spring studied more in order to be able successfully solve these problems. However, spring students had no opportunity to improve their performance and a smaller percentage of students actually demonstrated that they could solve these engineering economics problems.

Students in the spring performed better than the fall students in their best attempt for problems 7, 10, and 11. The difference in the success rate between spring and fall in the best attempt is not statistically significant for problems 7 and 11. Problem 7 was very difficult, and only 27% of students in the spring solved problem 7 correctly. Students in the fall also thought problem 7 was difficult, and it appears that many students in the fall strategically decided not to learn how to correctly solve the problem because they could still pass the testing module by correctly answering the other questions. Problem 11 was also fairly challenging and was part of testing module 7 in the fall. Testing module 7 which covered material at the end of the semester had the lowest passing rate because students had less time to improve their performance. Determining why students in the fall performed so badly on problem 10 compared to students in the spring is more challenging. One possible explanation is that the answer to the problem was a negative number, and many students in the fall entered a positive number, which was marked incorrectly by the online testing module.

More than half of the class in the fall semester responded in the survey that they would recommend the online testing modules in other engineering courses. Many students appreciated that they could receive instantaneous feedback (Chris Ricketts & Wilks, 2001), learn how to answer the questions, and the retake the testing modules. Such a method enables the students to learn from their mistakes without being anxious about their grades. These results correspond to the findings of the Numeracy and Statistics course to the first year Biology students where 88% of students liked having their points available instantly (Chris Ricketts & Wilks, 2001). The online testing modules allowed students to take the testing modules at their leisure at times they choose without sitting in a proctored room for an hour or two.

The overall performance of the students in fall was better than the students in the spring as the average grade in the fall was a B+ and the average grade in the spring was a B. More than half the students in the fall earned an A or an A- and were motivated to learn the additional material. However, 7% of the students in the fall failed the course, which may demonstrate that the stringent passing criteria can hamper the motivation of students to participate and learn from their mistakes. Implementing deadlines for testing modules could help in motivating students to complete the testing modules more quickly rather than solving all the testing modules towards the end of the semester and getting frustrated due to its repetitive structure. Although we did not specifically measure student strategies, we observed that students who were more proactive in taking testing modules frequently during the semester and who asked for help with questions that they answered incorrectly were more successful with the class.

A limitation of this study is that there is no direct comparison of whether students who are evaluated via repeated attempts of online tests retain course material better or worse than students who are evaluated via traditional methods. Future research can explore how much material students retain after a course with each type of assessment method. This new method of online assessment with multiple attempts could be beneficial for large classes and could especially prove to be fruitful for courses that are delivered online. Future research could analyze data on how much time students spend with the testing modules and when they take the testing modules in order to develop strategies may be best for succeeding with online testing modules. Studies could also be conducted that combine online testing with more conventional methods of assessment in engineering education to explore the effectiveness of such an approach.

Conclusion

Online testing modules could be one of the potential methods of conducting large engineering classes. It saves the instructors and the teaching assistants time, and more importantly it is a useful tool for distant-learning students. Students who feel anxious during in-class exams could benefit by online testing modules which would reduce their anxiety of writing test in a proctored setting. Students earned a better overall grade in the fall semester than the students who took in-class exams in spring, which shows that online tests with multiple attempts can help students earn better scores. Attempting a question with similar concepts multiple times on the testing modules helped in understanding the problem and improving through their mistakes. This could be because students practiced on the tests itself without worrying about the results. Hence, the authors feel that online testing modules could be an option for courses where student involvement and practice is expected.

Future research could be done to understand if the use of such online testing modules can retain knowledge for longer period than the traditional testing. The online testing modules did not consider referring to lecture notes, online content related to the course, or referring to previous test questions as cheating. Undoubtedly, some students did help each other answer questions while taking a testing module. This practice was prohibited, but we did not have the means to enforce this policy. In the future, research could be done to study if proctored settings allowing multiple attempts of exams help students learn better with minimum possibility of cheating. Moreover, it would be interesting to understand if cheating on online testing modules occurs more frequently than cheating on other types of assignments and in-class examinations.

References

Allen, I. E., & Seaman, J. (2008). Staying the course: Online education in the United States, 2008: *ERIC*.

Ammari, A. C., & Slama, J. B. H. (2006). The development of a remote laboratory for internet-based engineering education. *Journal of Asynchronous Learning Networks, 10*(4), 3-13.

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4-5), 514-527.

Charman, D., & Elmes, A. (1998). Formative assessment in a basic geographical statistics module. *Computer based assessment, 2,* 17-20.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE bulletin, 3,* 7.

Desouza, E., & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education, 14*(2), 121-134.

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education, 29*(5), 611-623.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Gates, A. I. (1917). Recitation as a factor in memorizing / by Arthur I. Gates. New York: New York : *Science Press*.

Gillet, D., Ngoc, A. V. N., & Rekik, Y. (2005). Collaborative Web-based experimentation in flexible engineering education.(Author Abstract). *IEEE Transactions on Education, 48*(4), 696. doi:10.1109/TE.2005.852592

Henson, A. B., Fridley, K. J., Pollock, D. G., & Brahler, C. J. (2002). Efficacy of Interactive Internet-Based Education in Structural Timber Design. *Journal of Engineering Education, 91*(4), 371-378. doi:10.1002/j.2168-9830.2002.tb00719.x

Jones, H. (1923). The effects of examination on the performance of learning. *Archives of Psychology, 10,* 1-70.

Kennedy, K., Nowak, S., Raghuraman, R., Thomas, J., & Davis, S. F. (2000). Academic dishonesty and distance learning: Student and faculty views. *College Student Journal, 34*(2).

Kian-Sam, H., Abang Ahmad, R., & Ming-Koon, K. (2003). Students' attitudes toward the use of the Internet for learning: A study at a university in Malaysia. *Journal of Educational Technology & Society, 6*(2), 45-49.

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education, 31*(3), 253-260.

Kingston, N., & Turner, N. (1984). Analysis of score change patterns of examinees repeating the Graduate Record Examinations General Test. ETS Research Report Series, 1984(1).

Leeming, F. C. (2002). The Exam-A-Day Procedure Improves Performance in Psychology Classes. *Teaching of Psychology, 29*(3), 210-212. doi:10.1207/S15328023TOP2903_06

Lyle, K. B., & Crawford, N. A. (2011). Retrieving Essential Material at the End of Lectures Improves Performance on Statistics Exams. *Teaching of Psychology, 38*(2), 94-97. doi:10.1177/0098628311401587

Maki, R. H., Maki, W. S., Patterson, M., & Whittaker, P. D. (2000). Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. *Behavior research methods, instruments, & computers, 32*(2), 230-239.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education, 31*(2), 199-218. doi:10.1080/03075070600572090

Passerini, K., & Granger, M. J. (2000). A developmental model for distance learning using the Internet. *Computers & Education, 34*(1), 1-15. doi:10.1016/S0360-1315(99)00024-X

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PloS one, 8*(11), e79774.

Ricketts, C., & Wilks, S. (2001). Is computer-based assessment good for students?

Ricketts, C., & Wilks, S. J. (2002). Improving Student Performance through Computer-Based Assessment: Insights from Recent Research. *Assessment & Evaluation in Higher Education, 27*(5), 475-479.

Ruffins, P. (2007). A real fear: it's more than stage fright. Math anxiety can derail academic or professional success but some scholars are working to help students get over it. *Diverse Issues in Higher Education.* Findarticle. com (online) http://findarticles. com/p/articles/mi_m0WMX/is_2_24/ai_n18744928.

Salzmann, C., Gillet, D., & Huguenin, P. (2000). Introduction to Real-time Control using LabVIEW (TM) with an Application to Distance Learning. *Int. J. Eng. Educ*, *16*(3), 255-272.

Sarason, S. B., & Mandler, G. (1952). Some correlates of test anxiety. *The Journal of Abnormal and Social Psychology*, *47*(4), 810.

Sarason, S. B., Mandler, G., & Craighill, P. G. (1952). The effect of differential instructions on anxiety and learning. *The Journal of Abnormal and Social Psychology, 47*(2S), 561.

Stowell, J. R., & Bennett, D. (2010). Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research, 42*(2), 161-171.

Terry, W. S. (2015). Learning and memory: Basic principles, processes, and procedures: *Psychology Press*.

Thalheimer, W. (2006). Spacing learning events over time: What the research says. Retrieved March, 21, 2007.

Tryon, G. S. (1980). The Measurement and Treatment of Test Anxiety. *Review of Educational Research, 50*(2), 343-372. doi:10.3102/00346543050002343

Vitasari, P., Wahab, M. N. A., Othman, A., Herawan, T., & Sinnadurai, S. K. (2010). The Relationship between Study Anxiety and Academic Performance among Engineering Students. *Procedia - Social and Behavioral Sciences, 8*(C), 490-497. doi:10.1016/j.sbspro.2010.12.067

White, R. J., & Hammer, C. A. (2000). Quiz-o-Matic: A free Web-based tool for construction of self-scoring on-line quizzes. *Behavior Research Methods, Instruments, & Computers, 32*(2), 250-253.

Willingham, D. T. (2004). Ask the Cognitive Scientist Practice Makes Perfect, But Only If You Practice Beyond the Point of Perfection. *American Educator*, *28*(1), 31-33.

Wolkowitz, A. A. (2011). Multiple attempts on a nursing admissions examination: Effects on the total score. *Journal of Nursing Education, 50*(9), 493-501.

Zacharis, N. Z. (2010). Innovative assessment for learning enhancement: Issues and practices. *Contemporary Issues in Education Research, 3*(1), 61-70.

CHAPTER 4. **CONCLUSION**

Online testing modules with multiple attempts has its pros and cons. It seems to be very beneficial for courses with large engineering class and distance learning courses. Answering such tests reduce anxiety of writing exams in proctored settings. Students get more chances to learn through their mistakes without worrying about the results of the exams. However, future research needs to be done on developing new methods to foolproof the testing method to avoid cheating by communicating while talking the test.

APPENDIX IRB EXEMPTION FORM

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date: 7/21/2017

To: Dr. Cameron MacKenzie
3029 Black Engr

From: Office for Responsible Research

Title: Engaging Students through Online Testing Modules for a High-Enrollment Engineering Economics Course

IRB ID: 17-351

Study Review Date: 7/21/2017

The project referenced above has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b) because it meets the following federal requirements for exemption:

- (1) Research conducted in established or commonly accepted education settings involving normal education practices, such as:
  - Research on regular and special education instructional strategies; or
  - Research on the effectiveness of, or the comparison among, instructional techniques, curricula, or classroom management methods.

- (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey or interview procedures with adults or observation of public behavior where
  - Information obtained is recorded in such a manner that human subjects cannot be identified directly or through identifiers linked to the subjects; or
  - Any disclosure of the human subjects' responses outside the research could not reasonably place the subject at risk of criminal or civil liability or be damaging to their financial standing, employability, or reputation.

The determination of exemption means that:

- **You do not need to submit an application for annual continuing review.**

- **You must carry out the research as described in the IRB application.** Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, changes in confidentiality measures, etc.), modifications that result in the inclusion of participants from vulnerable populations, and/or any change that may increase the risk or discomfort to participants. Changes to key personnel must also be approved. The purpose of review is to determine if the project still meets the federal criteria for exemption.

  Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

  Detailed information about requirements for submission of modifications can be found on the Exempt Study Modification Form. A Personnel Change Form may be submitted when the only modification involves changes in study staff. If it is determined that exemption is no longer warranted, then an Application for Approval of Research Involving Humans Form will need to be submitted and approved before proceeding with data collection.

Please note that you must submit all research involving human participants for review. **Only the IRB *or designees* may make the determination of exemption,** even if you conduct a study in the future that is exactly like this study.

Please be aware that **approval from other entities may also be needed.** For example, access to data from private records (e.g.

student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **An IRB determination of exemption in no way implies or guarantees that permission from these other entities *will* be granted.**

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.