Comparing Mastery-based Testing with Traditional Testing in Calculus II

Amanda Harsy[1] and Alyssa Hoofnagle[2]

[1] Lewis University
[2] Wittenberg University

Received: 15 November 2019; Accepted: 21 September 2020

Abstract

The authors present the results of quantitative data and student feedback from a two-year study of the effects of a new assessment model called "mastery-based testing." In mastery-based testing, students are given problems in which they can only receive full credit for the problem after they demonstrate mastery of the objective being assessed. This method strives to increase complete understanding of concepts through a growth-mindset approach to learning in addition to helping alleviate math anxiety. In this study, we compare the impact of mastery-based testing with traditional testing in six Calculus II classes. The data sources for the project included end-of-semester surveys, a final mastery assessment, and end-of-semester course grades. We found that mastery-based testing led to students feeling the assessments better reflected their content knowledge as well as higher end-of-semester course grades with fewer hours spent studying outside of class.

As instructors, some of the goals we have for our students include gaining a deep understanding of course content, strong problem-solving skills, and the confidence to tackle a problem until a solution is reached. All of these goals are centered on fostering what developmental psychologist, Carol Dweck, calls a "growth-mindset" approach towards learning, something that traditional points-based assessments often lack (Dweck, 2007). In an effort to encourage mathematics students to spend more time developing complete understanding of course material, mastery grading techniques like "Mastery-based Testing" are having a resurgence (Collins et al., 2018; Harsy, 2019; Heubach et al. , 2019; Mangum, 2019; Zimmerman, 2020). The structure of mastery-based testing may help reduce test anxiety and help motivate students to revisit old ideas that they have not fully understood until they are able to demonstrate mastery. Most authors like Collins et al. (2018), Harsy (2019), Heubach et al. (2019), Mangum (2019), and Zimmerman (2020) who have written about mastery-based testing have focused on its implementation and have not formally studied the impact of this assessment technique.

It is easy to get caught up in the enthusiasm when a new teaching method is introduced! It is especially exciting when the method supports some of the pedagogical goals one has as a teacher. But how do you know whether the new method is actually achieving those goals? When taking the plunge into the uncharted waters, how do you know whether the change will yield the pedagogical results you want? To better study the impact of mastery-based testing, we conducted a two year study comparing Calculus II courses taught by a single teacher. Students in these courses were given an end-of-semester survey which asked whether students felt that they fully mastered concepts, to what extent the students felt levels of test anxiety, whether they went back and studied past concepts, and if they felt that they better understood the material after studying the topics multiple times.

The following research questions guided this study:

1. **Does mastery-based testing impact how students approach studying?**
2. **Does mastery-based testing impact student reflection on their own knowledge?[1]**
3. **Does mastery-based testing impact end of semester content knowledge and grades?**

This paper reports the results of this study. We first outline some of the specifics about mastery-based testing (MBT). We then discuss our motivation for MBT and give the details of our experiment and describe our implementation of mastery-based testing in Calculus II classes. We report the results of our study and outline future research goals in the Results Section and Future Research Section.

Background of Mastery-Learning Techniques

In 2011, the National Center for Educational Achievement (NCEA, 2011) identified two characteristics of the highest performing schools from over 300 school districts. One was an alignment of the curriculum to the needs of the students to properly introduce, develop and master content, and the second was the assessment of concepts at each grade level as a prerequisite for advancement (Foshee et al, 2016). Foshee et al. (2016) interpreted this as a call for adaptive instruction and a mastery-based approach. In particular, teachers notice that students do not always learn material at the same pace and many experience varying levels of test anxiety when taking exams. Because of this many teachers have begun expressing concerns regarding the fairness and effectiveness of high stakes testing (Zimmerman & Dibenedetto, 2008).

Even before the NCEA's report, researchers were exploring other assessment methods. Carroll conjectured that any student can learn something as long as they have sufficient time. He also theorized that students have varying degrees of perseverance and ability to understand instruction, in addition to having varying opportunities for and quality of instruction (Zimmerman & Dibenedetto, 2008). In 1969, Bloom introduced mastery-learning methods in order to address Carroll's basic assumptions about

https://doi.org/10.20429/ijsotl.2020.140210

learning, assessment, and instruction (Bloom, 1968). Specifically, the term "mastery learning" is used to refer to variety of teaching techniques.

According to Slavin (1987), there are three forms of mastery learning: instruction, continuous progress, and learning for mastery. One, called the "Personalized System of Instructor" (PSI) or the "Keller Plan," allows students to take assessments over pre-established learning objectives as often as they wish (Keller, 1968). Usually this form of assessment has students working primarily at their own pace with self-instructional materials. This system is often popular for review courses or courses set up using online homework and exam platform systems like MyMathLab, ALEKS, or WebAssign. A second form, "Continuous Progress," allows students to work on concept units at their own rate but employs mastery-based grading criteria. The teacher will need to provide supplemental activities for students who do not master material during their first attempts (Cohen, 1977; Slavin, 1987). The third variation of mastery learning is called "Group-based Mastery Learning" or "Learning for Mastery" (Block & Anderson, 1975). For this version of mastery-based learning, the teacher will instruct the class as a whole and give exams in which the students need to reach a predetermined percentage on the assessment in order to "pass." Students who do not pass on the first attempt will need to receive support through tutoring and must continue to retest until they meet the benchmark. Since this form of mastery learning isn't completely self-paced, the class will move on to new topics regardless of whether all of the students have passed the assessments.

Several educational studies have already reported on the effectiveness of mastery-learning approaches for students and teachers (Kulik et al., 1990; Zimmerman & Dibenedetto, 2008). In their meta-analysis, Kulik et al. (1990) found positive effects of mastery-learning approaches when analyzing results of final exam performances in 93% of their studies (62 out of 67 studies of the programs they studied). In particular, they found that programs which used some type of mastery learning (if even only slightly) had positive effects on student attitudes regarding the content of the course and instructional methods. They found an even stronger positive effect for mastery-learning students after an 8-week delay, which they felt supported the hypothesis that mastery-learning had positive effects on retention of concepts (Kulik et al., 1990). Additionally, Kulik, et al. found more consistency in learning of course objectives in mastery-learning classrooms. That is, they found less variation between students in mastery-learning based classrooms than in the control classrooms. Furthermore, they noticed that students in mastery-learning courses "rated the quality of their instruction and their attitude toward the subject matter more favorably than students in traditional classes" (Kulik et al., 1990). In their meta-analysis, Guskey and Gates (1986) found that mastery-learning in the classroom seemed to reduce the correction time students needed as students continued through the objectives of the course. They also found motivation boosts for both the student and the instructors in mastery-learning environments. In particular, they noticed that students in mastery-learning environments developed better positive attitudes about their ability to succeed in the course in addition to spending more time engaged in the learning process than students in the control classrooms (Guskey & Gates, 1986). Moreover, they found that teachers who used mastery-learning approaches in their classes had higher expectations for the students in their classes, felt more personal responsibility towards their student's ability to succeed in learning the course objectives, and also had better attitudes about teaching in general than their traditional counterparts (Guskey & Gates, 1986).

The studies from Guskey and Gates (1986) and Kulik, et al. (1990) seem to reflect Bloom's (1968) belief that there would be benefits for students in mastery-learning teaching environments. Furthermore, the results of these studies seem to support that mastery-learning can help students improve their "self-efficacy," the belief they have about their ability or lack of ability to perform particular tasks or actions (Foshee et al., 2016). A strong sense of self-efficacy has many benefits for students. It has been shown to increase personal motivation, improve feelings of competency and self-worth along with being a positive predictor of performance (Foshee et al., 2016). Often when a student has a poor sense of self-efficacy about a certain task, they tend to avoid doing that task (Foshee et al., 2016). Mastery-learning teaching approaches help to counteract a student's desire to avoid difficult tasks by helping to increase students' resilience and motivation to continue working on course objectives and helped them develop a positive self-efficacy (Guskey & Gates, 1986; Kulik et al., 1990). Furthermore, mastery-learning approaches seem to help students think about mistakes as learning opportunities rather than indicators of low ability, since students are able to recover and learn from past mistakes on concepts (Boaler, 2013). Therefore, mastery-learning seems to foster what developmental psychologist, Carol Dweck, calls a "growth-mindset" approach towards learning since, through effort and practice, students who work hard and learn from their mistakes can persevere through course concepts. (Dweck, 2007, 2013).

What is Mastery-Based Testing?

The mastery-learning scheme used in this study follows Collins, et al's (2018) assessment technique called "mastery-based testing" (MBT). This implementation of mastery-learning adheres to the common essential features of mastery-based testing: "clear course concepts, credit only for mastery, and multiple attempts to display mastery" (Collins et al., 2018). Thus, MBT is a blend of the mastery-learning approaches Slavin (1987) labeled "Continuous Progress'" and "Learning for Mastery." Students are graded using a mastery criteria (not percentage-based) and have multiple opportunities to demonstrate this mastery and receive help on the material. Unlike the models in which students work through the material completely at their own pace, the class is led by an instructor which means the class moves on regardless of whether students have mastered previous material. In this implementation, students can continue to work on mastering past material as they continue learning new concepts or objectives.

In MBT, students have multiple opportunities to demonstrate learning of concepts, but only receive credit when they display a "mastery" level of understanding. If a student does not master the material on the first try, re-testing opportunities are available on future exams, quizzes, or testing weeks. To implement MBT, an instructor will first need to break up their semester-long content into a certain number of course topics or concepts (motivated by the course student learning outcomes). Most MBT courses have 14-18 concepts and students can add to their exam grade by mastering any of those concepts. In some courses, it may make sense to let the students choose which concepts they want to master. But if an instructor believes that there are some concepts that students need to have grasped at the completion of the course, one common modification for MBT is to use "core concepts" (Collins et al., 2018). Students are then able to retest concepts naturally by the structure of the exams. One common break down of an exam structure for a class with 16 concepts could be as follows: exam 1 could start with four questions over the first four concepts. Exam 2 could then have 8 questions - revised versions of the four from exam 1 and four new questions. Exam 3 could have 12 questions- revised versions of the eight from exam 2 and four new questions. Exam 4 will consist of versions of the twelve from exam 3 and four new questions on the remaining concepts. The final exam usually has no new questions and contains only new versions of the sixteen concepts. Note that although the questions from exam 2 over the concepts in exam 1 are not the same, they should test the same concept. Once a concept is mastered, the student receives full credit and need never attempt the question again (even though it will appear on the test sheet). Thus, if a student has mastered all of the concepts before the final exam, they would not need to take the final exam. A student's final test grade is determined by the number of concepts they have mastered by the end of the semester.

Key components of successfully implementing mastery-based testing include providing multiple opportunities to demonstrate mastery, giving timely feedback on these attempts, and an objective rubric for mastery grading (Collins et al., 2018). It is important to allow multiple opportunities to demonstrate mastery of each learning outcome in a way that once a concept is mastered, previous failed attempts will not adversely affect a student's grade. There are many ways to allow for multiple attempts and provide timely feedback on these attempts. Some instructors use weekly quizzes, office hours, or testing weeks for these retesting opportunities (Collins et al., 2018). It is important for instructors to create rich mastery-concept questions which test the same concepts but can be changed enough to prevent students from mimicking/memorizing patterns of answers. Finally, instructors need to determine ahead of time what characteristics a student's solution must include to constitute mastery. It is helpful for the instructor to take some time explaining the mastery grading procedures to students, outlining what type of work is expected to earn mastery. Many students are initially worried that they need to provide a solution that is 100% correct in order to earn mastery; thus, a conversation at the start of the course is beneficial in clearing up any misconceptions. Most instructors will give a mastery grade when they believe the student has demonstrated understanding of the concept (with maybe some minor errors irrelevant to the concept itself) and would not necessarily benefit from revisiting the concept again.

Motivation for Implementation of Mastery-based Testing

One of the goals of mastery-based testing is to motivate students to revisit old ideas that they have not fully understood and recognize that mistakes are opportunities for learning and growth (Boaler, 2013). MBT provides students opportunities to master and fully understand previous concepts which will aid in their learning of new material. In this way, MBT supports a growth-mindset towards learning, since students who work hard and learn from their mistakes can persevere through the material (Dweck, 2007, 2013). By supporting the idea that through effort and practice, students can develop their learning abilities, MBT allows students to change how they approach mistakes. First, it reduces the cost of mistakes in assessment opportunities by giving students a chance to learn and re-evaluate concepts with no penalty. Second, students are incentivized to keep working on a concept until they can show mastery of that concept. In this way, MBT forces students to thoroughly understand a mathematical concept at some point during the semester. Additionally, mastery-based testing may help to reduce test anxiety, since students will have multiple times to attempt concepts.

Some assessment techniques have already aimed to counteract a "fixed" learning culture. While these techniques support our efforts, they are not the same as mastery-based testing. For example, some teachers use standards-based grading or no-points grading (Brilleslyper et al., 2012). In Beatty (2013), students earned points per standard or objective for the course not through exam points. In this course, normal scores were used, not mastery grading, but like MBT these scores relate to a particular skill or objective for the course. Also, Studman (1984) used a version of mastery learning, in which he identified a set of objectives which his students needed to master to pass the course. Like in MBT, these students were allowed multiple attempts to show mastery, but unlike MBT, testing could occur at any time rather than during regular tests or quizzes. Also, in this approach, mastery does not entail full conceptual understanding, but instead a student just needs to show a general knowledge of the content, which is similar to earning a C. All of the above techniques take a different approach than MBT, making MBT a new and innovative assessment strategy.

METHODS

This study was conducted at a four-year, private, primarily teaching-focused university whose enrollment is approximately 6,500 with an undergraduate population of around 4,000. The university services a 34% minority population and is an Emerging Hispanic-Serving Institution. Thirty percent of undergraduates are first generation college students and thirty-one percent of undergraduates receive Pell grant which are awarded to low-income students. This study compared MBT and traditional assessment in six Calculus II courses: two in Spring 2015, one in Fall 2015, two in Spring 2016, one in Fall 2016. The 96 students enrolled in Calculus II courses during this time period reflects the following demographics: 69% (n=66) were male while 31% (n=30) were female; 40% (n=39) were freshmen, 25% (n=24) were sophomores, 17% (n=16) were juniors, and 18% (n=17) were seniors; 90% (n=86) were STEM (biology, chemistry, physics, computer science, mathematics) or computer engineering majors. See Table 1 for a description of the participant sample.

Of the 96 students surveyed, 53 were in MBT classes and 43 were in traditional classes each from two spring classes and one fall. The average class size was 18. All classes were taught by the same professor and traditional grades were assigned for homework and bi-weekly quizzes over online homework. The only difference was the assessment method for exams. In both courses, quizzes and homework counted as 15% and 20% of their grade, respectively. Exams in MBT classes counted towards 65% of the final grade. In traditionally assessed classes, exams were worth 45% of the final grade and the cumulative final counted for 20% of the grade. The lowest regular exam score was replaced by the cumulative final exam score if it improved the score. In

Table 1. *Characteristics of Participant Sample (n=96)*	
Type of Testing	n (%)
Mastery-Based Testing	53 (55%)
Traditional Testing	43 (45%)
Gender	
Male	66 (69%)
Female	30 (31%)
Year in School	
Freshman	39 (40%)
Sophomore	24 (25%)
Junior	16 (17%)
Senior	17 (18%)
Major	
STEM Fields	86 (90%)
non-STEM Fields	10 (10%)

addition to the goals of the research in Collins et al. (2018), we used a 14-question final mastery assessment to examine whether students were able to retain concepts better with MBT. This final mastery assessment was given on the last day of class and graded on a 1-point scale without partial credit. This grading scale is harsher than mastery, since any wrong answer yielded no credit. Students did not know which subset of concepts were going to be on the final mastery assessment but were told it was cumulative. To motivate students to work hard on the final mastery assessment, we also counted this assessment as a quiz grade. For MBT, we allowed students to use the final mastery assessment to master any concept on the assessment. For traditional students, we used a scale of their score on this final mastery assessment for the bonus questions on the cumulative final. At the end of the semester, students were given this common, final mastery assessment along with anonymous surveys about their thoughts on the course. The survey was approved by our Institutional Review Board and is included in the Appendix and was the same used in Collins et al. (2018).

Implementation of MBT

As mentioned earlier, there are many ways of implementing mastery-based testing. For the Calculus II courses in this study, we used 16 total concepts and identified seven of these concepts as core concepts. We felt all Calculus II students should master these seven core concepts in order to be successful in Calculus III. Therefore, students had to master the core concepts in order to earn at least a C for their exam grade. Any additional concepts the student mastered increased their exam grade.

Students in the mastery-based testing classes during this study had four in-class exams and four testing weeks in-between the exams. These testing weeks allowed students to retest past concepts once during the week. For example, a student could retest concepts 2 and 3 on Monday and concept 4 on Tuesday, but they could not retest concept 2 a second time during the week. These retesting opportunities were done during office hours or proctored at the university study tables.

Implementation of Traditional Testing

Students in the traditional testing were taught the material in the same way as the MBT students. They had the same instructor, same homework assignments, similar bi-weekly quizzes (with problems taken from their online homework system), the same lecture note guides (a 200+ workbook the course worked through), and the same final mastery assessment which was given to MBT students. Traditionally assessed students were given 4 exams which were broken down in the same way as the MBT exams and a cumulative final exam, all of which were graded with a traditional points-based, partial credit system. This final exam would replace the lowest of the 4 regular exams if the final exam score was higher than the regular exam score. Students in the traditional course also had an optional review day before the final exam.

DATA SOURCES AND COLLECTION

All students enrolled in Calculus II during the 2015 and 2016 fall and spring semesters were invited to participate in the data collection; however, a student could refuse to provide feedback on the end-of-semester survey and/or not include his or her final grades. All feedback was kept anonymous and any grades were kept confidential.

Analysis

Data collected from student surveys were analyzed directly to determine whether the responses and final grades were dependent on the assessment method (traditional or MBT) used in the course. All student responses available were included, and students could opt out of the data collection at the start of the semester. Since the same surveys were administered each semester, we did not need to use normalized scores prior to analysis. For each question on the survey with Likert responses, we set 1 = Strongly Disagree to 4 = Strongly Agree and used a Welch's *t*-test with a significance level of $\alpha=0.05$ to determine if the average response differed between traditional assessment and MBT. For the survey question regarding the number of hours per week students spent on the course, we used an ordinal logistic regression test due to the ordered categorical responses. We also collected data on the students' end of the semester course grades with an A = 4.0, a B = 3.0, etc., their grade on the final mastery assessment, and the number of concepts they felt they had mastered. Again, students could choose not to be included, and we did not need to normalize due to the consistent data collection across semesters. Welch's *t*-tests using a significance level of $\alpha=0.05$ were used to assess if the average end-of-semester grade, the average grade on the final mastery assessment, and the average number of concepts mastered differed between the two assessment methods.

For each survey question, we provide a graphical and tabular representation of the percentages of students' Likert ratings on the question in Appendix B.

RESULTS

The results are grouped according to the three main research questions stated in the Introduction.

Research Question 1: Does mastery-based testing impact how students approach studying?

The researchers used three survey questions to assess whether mastery-based testing had an impact on how students approached studying in Calculus II. These questions are:

- Studying for exams in this course helped me to learn the material.
- I relied mostly on memorizing solutions to earlier problems to prepare for in-class assessments.
- How many hours per week did you spend on this course outside of class time?

While we note that students may not be the best at assessing their own learning (Tousignant et al., 2002; Falchikov, et al., 1989; Brown et al., 2015), we did want their feedback on their own reflection on how they learned. Although the questions about studying for exams and memorizing solutions may not be unbiased, the concerns brought up by Collins et al (2018) and Harsy (2019) prompted us to explore this. The mean response to the survey question "Studying for exams in this course helped me to learn the material" was 3.538 for MBT students (SD = 0.536) and was 3.535 for traditional students (SD = 0.585). The mean response to the survey question "I relied mostly on memorizing solutions to earlier problems to prepare for in-class assessments" was 2.019 for MBT students (SD = 0.713) and was 2.233 for traditional students (SD = 0.858). Possible responses for the question, "How many hours per week did you spend on this course outside of class time?" were categorical groups, such as "3-5 hours" or "12-14 hours." In the MBT group, 37% (n=20) of students reported spending 0-2 hours per week, 45% (n=24) reported spending 3-5 hours per week, and 16% (n=8) reported spending 6 or more hours per week (with 1 student who did not respond). In the traditional group, 14% (n=6) of students reported spending 0-2 hours per week, 46% (n=20) reported spending 3-5 hours per week, and 40% (n=17) reported spending 6 or more hours per week More information about the sample responses to these three survey questions can be found in Appendix B.

The researchers found no significant difference in testing method when students assessed whether their studying helped them learn the course material (t = 0.03; p = 0.976). See Table 2 for additional details of the test. Most students in either group agreed or strongly agreed to this survey question (96% of MBT students and 95% of traditional students). This supports the conclusion that mastery-based testing does not influence students' belief that studying helped them learn the material.

There was also no significant difference in testing method when students responded to whether they relied on memorizing solutions to earlier problems to prepare for upcoming in-class assessments (t = –1.29; p = 0.2). See Table 2 for additional details of the test. This result is interesting, since it is important for instructors to create rich mastery-concept questions to prevent students from mimicking patterns of solutions. In fact, the researchers note that none of the students in the mastery-based testing group strongly agreed that they were using memorization to prepare for their assessments.

When conducting the ordinal logistic regression test, the researchers did find a significant difference in testing method when students reported how much time they spent outside of class (Z = 3.07; p = 0.002). The Pearson Chi-square Goodness-of-Fit test was not significant, as desired in the ordinal logistic regression analysis (chi-square = 6.01; df = 4; p = 0.198). Overall, mastery students do not report studying as much as traditionally assessed students, with a parameter estimate of 1.24. The researchers speculate that this is because mastery students would focus on fewer concepts at a time rather than cramming for a huge midterm exam.

In summary, students in both groups reported similar study and memorization patterns. However, the mastery-based testing students self-reported spending less time outside of class studying - perhaps due to the fact that some students choose to learn at their own pace rather than being dictated by the instructor's pace. The researchers conclude that the students' study habits were not significantly impacted by using mastery-based testing.

Table 2. *Results of Welch's t-Tests Comparing Study Habits between Testing Types*

	Mean (MBT Group)	Mean (Trad. Group)	SE	df	t	p-value
Studying helped learn course content	3.538	3.535	0.117	86	0.03	.976
Relied on memorizing solutions to prepare	2.019	2.233	0.162	81	-1.29	.200

Research Question 2: Does mastery-based testing impact student reflection on their own knowledge?

In order to assess this research question, the researchers used two survey questions along with the number of concepts the students self-reported mastering. The survey questions are:

- The results of my in-class assessments accurately reflect my knowledge.
- The in-class assessments deepened my understanding of the ideas in this course.

The mean response to the survey question "The results of my in-class assessments accurately reflect my knowledge" was 3.321 for MBT students (SD = 0.576) and was 3.049 for traditional students (SD = 0.764). The mean response to the survey question "The in-class assessments deepened my understanding of the ideas in this course" was 3.472 for MBT students (SD = 0.536) and was 3.190 for traditional students (SD = 0.663). More information about the sample responses to these survey questions can be found in Appendix B.

Upon completing a two-tailed test, there was no significant difference in testing method when students assessed whether the in-class assessments accurately reflected their knowledge (t=1.88; p=0.064). However, the researchers note that the p-value is close to the 0.05 significance level, indicating some evidence of a difference in testing method. See Table 3 for more details of the test. The difference is seen in the sample with a 20% difference in the two sampled groups who agreed or strongly agreed (MBT: 94%; traditional: 74%) and a 15% difference in the two sampled groups who disagreed or strongly disagreed (MBT: 6%; traditional: 21%).

There was a significant difference in testing method when students responded to whether the in-class assessments deepened their understanding of the course content (t=2.21; p=0.03). See Table 3 for more details of the test. The major difference is that mastery-based testing students are more likely to believe the assessments deepen their understanding of ideas more than traditional students, with 98% (n=52) of mastery-based testing students agreeing or strongly agreeing compared to 84% (n=36) of traditional students. In fact, only one of the mastery students disagreed with this statement, and none strongly disagreed.

An additional question on the survey included different concepts covered in the Calculus II course. These went beyond the 16 mastery concepts and included 40 concepts covered

Table 3. *Results of Welch's t-Tests Comparing Student Reflection of Knowledge between Testing Types*

	Mean (MBT Group)	Mean (Trad. Group)	SE	df	t	p-value
Accurately reflect knowledge	3.321	3.049	0.145	72	1.88	0.064
Deepened understanding	3.472	3.190	0.127	78	2.21	0.030*

*p<0.05, two-tailed.

throughout the semester. The 40 concepts are listed on the survey in Appendix A. The average number of concepts students self-reported mastering in the mastery-based testing group was 22.794 concepts (SD = 8.710), while the average in the traditional testing group was 22.163 concepts (SD = 8.529). Using a Welch's *t*-test, the researchers found no significant difference in the average number of concepts mastered between these two groups (t=0.40; p=0.694). See Table 4 for additional details of the test.

Overall, there is a significant difference in student opinion of the assessment method when it pertains to the content knowledge in the course, but students generally report understanding the same number of concepts for both mastery-based and traditional testing. The researchers conclude that mastery-based students leave the course with more positive feelings that the testing method accurately assesses and deepens their knowledge compared to traditionally tested students.

Research Question 3: Does mastery-based testing impact end of semester content knowledge and grades?

Finally, the researchers investigated whether mastery-based testing impacted their score on a final mastery assessment as well as course grades at the end of the semester. While the instructor assigned the full spectrum of +/- grades, the researchers groups grades by letter, A, B, C, D, and "Not Passing" (F's and Withdrawals) and assigned grade points (eg. A = 4.0, B = 3.0, etc.) in order to conduct the statistical analysis. The mean course grade point for MBT students was 2.918 (SD = 1.324) and the mean course grade point for traditional students was 2.537 (SD = 1.234).

At the conclusion of the semester, a final mastery assessment testing the course concepts was given to all students and graded on a binary scale: 100% correct (1 point) or incorrect (0 points). The scores for each student were compiled and recorded. The average score on this final mastery assessment for students in the mastery-based testing group was 5.037 (SD = 2.742), while the average in the traditional testing group was 4.703 (SD = 3.593). Using a Welch's *t*-test, the researchers found no significant difference in the average score on the final mastery assessment between the two groups (t=0.48; p=0.636). More details of the test are seen in Table 4.

However, there was a significant difference in end-of-semester course grades between the two methods, with a higher percentage of students earning A's with mastery-based testing (53%, n = 32) compared to traditional assessments (29%, n=14; t=1.68; p=0.048). See Table 4 for additional details of the test. Anecdotally, the researchers also observed that mastery-based testing does not seem to have much effect on the students who will get A's or the students who would most likely not pass the course regardless of the testing method used. However, mastery-based testing helps the hard-working B student by giving him or her a chance to work hard and earn an A. On the other side of the spectrum, there seems to be a higher percentage of students who do not pass when mastery-based testing is used (11% of MBT students and 8% of traditional students). The researchers believe this is due to the fact that when a student does not master concepts by the end of the semester, it causes the course grade to be very low. In a traditionally graded class, that same student would still not understand many concepts, but may be able to earn a D through partial credit on exams. We noticed that while both sets of students demonstrated similar knowledge of content (as demonstrated by the final mastery assessment), MBT classes had a more bimodal grade distribution, which better reflects students' actual content knowledge.

Table 4. *Results of Welch's t-Tests Comparing Content Retention and End-of-Semester Grade Points between Testing Types*

	Mean (MBT Group)	Mean (Trad. Group)	SE	df	t	p-value
Number of concepts mastered	22.794	22.162	1.588	33	0.40	0.694
Score on final mastery assessment	5.037	4.703	0.699	36	0.48	0.636
End-of-semester grade point	2.967	2.551	0.248	106	1.68	0.048†

†p<0.05, one-tailed, in favor of MBT.

Student Comments

We also collected feedback from the students in both classes. In general, MBT students were very receptive of the MBT assessment method. Some of the themes in the comments were appreciating extra chances to practice and improve their understanding of concepts, feeling in control of improving their grades through effort, and believing MBT forced them to gain complete understanding of concepts from the course.

One student said,

> [I] did prefer the mastery testing system as opposed to the normal testing system, it allows students to prove that they know the material the next time around if they happen to have an off day. Mastery also forces students to know the material throughout the semester, as it is always on the next test if they get it wrong.

Another student stated in his course evaluations:

> An experimental 'mastery' based testing system was used for my Calculus II class. I was skeptical of it at first, but it encouraged me to revisit topics I had previously failed to understand and allowed me to improve my grade from what it could have been with additional work and effort. It also allowed me to make up a weeks' worth of material missed due to unexpected illness. Going forward, I would definitely recommend this system be used for math classes in the future.

One student reported:

> I liked the idea of Mastery Based Testing. It was nice to get a second chance to understand a problem as students could only achieve a good grade through complete understanding of a problem. I believe that math revolves around getting enough practice, and sometimes coming back to a problem later on with more practice helps to understand it a little better.

Some students also mentioned that it helped to decrease their test anxiety. One shared, "I also enjoyed the mastery exams because they made exams much less stressful. I appreciated the ability to try concepts again, even if I did not do so well the first time." One survey question asked students whether they were anxious about exams. While this question alone does not show that mastery-based testing reduces anxiety, the researchers found that 19% (n=10) of MBT students strongly agreed compared to 37% (n=14) of traditional students who strongly agreed. Also, 21%

(n=11) of MBT students strongly disagreed with the statement whereas no traditional students strongly disagreed. This demonstrates that, at least in this sample, there were notable differences in the extreme responses related to test anxiety.

Even though students seemed to appreciate the chance to reexamine course concepts, one of the common complaints about MBT was the increased workload and effort they had to put in if they did not make consistent progress toward mastery of the concepts. One student said, "Mastery based testing is a great way to learn because it forces the student to completely understand and master a concept; however, the downside is that the test workload increases as more concepts are not mastered. The 'testing weeks' were a great supplement to mastering yet unmastered concepts."

A few students said that the extra opportunities encouraged them to procrastinate. One such student shared "I do not think the mastery-based testing is good. From a student standpoint it is good because I have multiple opportunities to master concepts, and if a test happens to fall on a day I have big assignments or tests in other classes I could focus a little more on those and just retest on a different day. With that being said I do not think I really learned much in the class, I would study enough to master the concept but I do not think I learned it. I do not think there should of been so many testing opportunities, if I do not study enough to pass the concepts the first time then it is my fault if come test three I have too many concepts and not enough class time to take the test." Despite some complaints, the majority of the feedback from students was positive and they requested that this assessment be used in other math courses.

CONCLUSION AND FUTURE WORK

This initial two-year study on the use of mastery-based testing found that there was no statistically significant difference between MBT students and traditional students on the number of concepts mastered on the end-of-semester final mastery assessment and number of concepts students reported they felt they had mastered. With regards to the survey, there was no difference between student responses to the following questions:

- Studying for exams in this course helped me to learn the material.
- I relied mostly on memorizing solutions to earlier problems to prepare for in-class assessments.
- The results of my in-class assessments accurately reflect my knowledge.

We did find a difference in a few of the survey responses. MBT students more strongly agree that the in-class assessments deepened their understanding of the ideas in the Calculus II course, and MBT students also reported feeling less anxious before exams. Finally, MBT students reported studying fewer hours each week than their traditional student counterparts and more MBT students earned higher final grades than traditional students. The researchers believe this is a primary reason to adopt mastery-based testing over traditional grading.

From this initial study, more questions came to light specifically with regards to measuring growth mindset and the nature of the anxiety that may be felt in the class. We have commenced a new two-year study which continues this work. In addition to the questions from the original study, we have included more in-depth questions about test anxiety and how much they examined and reflected on past material in addition to questions about growth mindset. We ask questions about their mindset on learning by asking them to rate the extent to which they agree with statements such as "You have a certain amount of math intelligence, and you can't really do much to change it." We also now administer a pre-survey, mid-survey, and post-survey to better track the growth mindset and anxiety levels throughout the course, and we also give an end-of-the-semester final mastery assessment. Finally, we are also expanding the scope of the project and are exploring how mastery-based testing affects student performance and test anxiety in an introductory statistics course.

ACKNOWLEDGEMENTS

The work was supported by a Lewis University Scholarship of Teaching and Learning Grant.

NOTES

1. We note that students may not be the best at assessing their own knowledge and learning (Tousignant et al. (2002), Falchikov et al. (1989)).

REFERENCES

Beatty, I. D. (2013). Standards-based grading in introductory university physics. *Journal of the Scholarship of Teaching and Learning*,13(2), 1–22.

Block, J. H., & Anderson, L.W. (1975). *Mastery learning in classroom instruction*. Macmillan Pub Co.

Bloom, B. S. (1968). Learning for mastery. instruction and curriculum. regional education laboratory for the carolinas and virginia, topical papers and reprints, number 1. *Evaluation comment*,1 (2), n2.

Boaler, J. (2013). Ability and mathematics: The mindset revolution that is reshaping education. In *Forum* (Vol. 55, pp. 143–152).

Brilleslyper, M., Ghrist, M., Holcomb, T., Schaubroeck, B., Warner, B., & Williams, S.(2012). What's the point? the benefits of grading without points. *Primus*, 22(5), 411–427.

Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. Assessment in Education: Principles, Policy & Practice, 22(4), 444-457.

Cohen, S. A. (1977). Instructional systems in reading: A report of the effects of a curriculum design based on a systems model. *Literacy Research and Instruction*,16(3), 158–171.

Collins, J. B., Ramsay, A. H., Hart, J., Haymaker, K. A., Hoofnagle, A. M., Janssen, M.,. OShaughnessy, J. (2018). Mastery-based testing in undergraduate mathematics courses. *Primus*, 29(5), 441-460. doi: 10.1080/10511970.2018.1488317

Dweck, C. S. (2007). Boosting achievement with messages that motivate. *Education Canada*, 47(2), 6–10.

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development.* Psychology press.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. Review of educational research, 59(4), 395-430.

Foshee, C. M., Elliott, S. N., & Atkinson, R. K. (2016). Technology-enhanced learning in college mathematics remediation. *British Journal of Educational Technology*, 47(5),893–905

Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery earning in elementary and secondary classrooms. *Educational leadership*, 43(8), 73.

Harsy, A. (2019). Variations in Mastery-based Testing. *Primus*, 30(8-10), 849-868. doi: 10.1080/10511970.2019.1709588

Heubach, S., & Krinsky, S. (2019). Implementing Mastery-Based Grading at Scale in Introductory Statistics. *Primus*, 30(8-10), 1054-1070. doi: 10.1080/10511970.2019.1700576.

Keller, F. S. (1968). "good-bye, teacher." 1. *Journal of applied behavior analysis*, 1(1), 79–89.

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of educational research*, 60(2), 265–299.

Mangum, A. (2019). Implementation of Mastery-Based Testing in Calculus. *Primus*, 30(8-10), 869-884. doi: 10.1080/10511970.2019.1709587.

Slavin, R. E. (1987). Mastery learning reconsidered. *Review of educational research*, 57(2), 175–213.

Studman, C. (1984). A method of applying mastery learning to moderately large classes. *International Journal of Mathematical Education in Science and Technology* ,15(1), 95–100.

Tousignant, M., & DesMarchais, J. E. (2002). Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: a correlation study. Advances in Health Sciences Education, 7(1), 19-27.

Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools*, 45(3), 206–216

Zimmerman, J. K. (2020). Implementing Standards-Based Grading in Large Courses Across Multiple Sections. *Primus*, 30(8-10), 1040-1053. doi: 10.1080/10511970.2020.1733149.

APPENDIX A

Assessment Questionnaire

The results of this questionnaire will be anonymous. I appreciate your honest feedback.

Instructions: Mark the response that most accurately reflects your opinion.

1. **The assessments in this course test our understanding of key concepts.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
2. **Studying for the exams in this course helped me to learn the material.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
3. **The results of my in-class assessments accurately reflect my knowledge.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
4. **The in-class assessments deepened my understanding of the ideas in this course.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
5. **I was anxious before the exams in this course.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
6. **During the course of the semester, my anxiety level**
   □ increased □ decreased □ stayed the same □ I had no anxiety □ no response
7. **I relied mostly on memorizing solutions to earlier problems to prepare for in-class assessments.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
8. **Thoughout the semester, I often revisited old ideas that I hadn't fully understood.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
9. **I have retained past concepts and ideas learned earlier in the course.**
   □ strongly agree □ agree □ disagree □ strongly disagree □ no response
10. **I have a good understanding on how material learned earlier in the semester relates to material learned later in the semester.**
    □ strongly agree □ agree □ disagree □ strongly disagree □ no response
11. **I feel prepared to approach a wide range of problems in Calculus II.**
    □ strongly agree □ agree □ disagree □ strongly disagree □ no response
12. **How many hours per week did you spend on this course outside of class time?**
    □ 0-2 hours □ 3-5 hours □ 6-8 hours □ 9-11 hours □ 12-14 hours □ more than 14 hours
13. **Which concepts from this course do you feel that you have mastered? Mark all that apply. (2 sides)**
    - □ One-to-one Functions
    - □ Inverse Functions
    - □ Exponential Functions
    - □ Logarithmic Functions
    - □ Derivatives of Exponential and Logarithmic Functions
    - □ Integrating with Exponential and Logarithmic Functions
    - □ Applying Exponential and Logarithmic Functions
    - □ Derivatives of Inverse Trig Functions (arcsine, arccosine, etc.)
    - □ Derivatives of Hyperbolic Functions ($\sinh x$, $\cosh x$, $\tanh x$)
    - □ Antiderivatives of Hyperbolic Functions ($\sinh x$, $\cosh x$, $\tanh x$)
    - □ Derivatives of Inverse Hyperbolic Functions ($\sinh^{-1} x$, $\cosh^{-1} x$, $\tanh^{-1} x$)
    - □ Integrating Using Inverse Hyperbolic Functions ($\sinh^{-1} x$, $\cosh^{-1} x$, $\tanh^{-1} x$)
    - □ Integrating Using Substitution
    - □ Computing Arc Length
    - □ Computing Surface Area
    - □ Computing the Area under Curves

- ☐ Shell Method of Computing the Volume of a Solid
- ☐ Disk/Washer Method of Computing the Volume of a Solid
- ☐ Computing Work Using Integration
- ☐ Computing The Center of Mass/Centroid Using Integration
- ☐ L'Hopital's Rule
- ☐ Integration by Parts
- ☐ Methods for Integrating Higher Powers of Trig Functions
- ☐ Inverse Trig Substitution Technique for Integration
- ☐ Partial Fraction Decomposition Technique for Integration
- ☐ Trapezoid Rule For Approximating an Integral
- ☐ Simpson's Rule For Approximating an Integral
- ☐ Calculating Improper Integrals
- ☐ Sequences
- ☐ Series
- ☐ Comparison Tests to Determine Series Convergence/Divergence
- ☐ Alternating Series Test
- ☐ Absolute VS Conditional Convergence
- ☐ P-Test for Series
- ☐ Integral Test to Determine Series Convergence/Divergence
- ☐ Geometric Series
- ☐ Ratio/Root Tests to Determine Series Convergence/Divergence
- ☐ Using Sequences of Partial Sums
- ☐ Power Series
- ☐ Taylor Series
- ☐ Taylor Series Approximation

14. Which study methods did you use? Mark all that apply.

- ☐ Practicing additional problems from the book
- ☐ Completing problems on review materials
- ☐ Reading the textbook
- ☐ Reading my notes
- ☐ Redoing homework problems
- ☐ Trying examples from the textbook or notes on my own
- ☐ Group studying
- ☐ Discussions with my instructor
- ☐ Recopying notes from class
- ☐ Watching videos online
- ☐ Math Study Tables
- ☐ Private Tutor
- ☐ LARC Tutoring
- ☐ Other: ______________________________

15. Additional comments regarding assessment in this course:

____

____

____

____

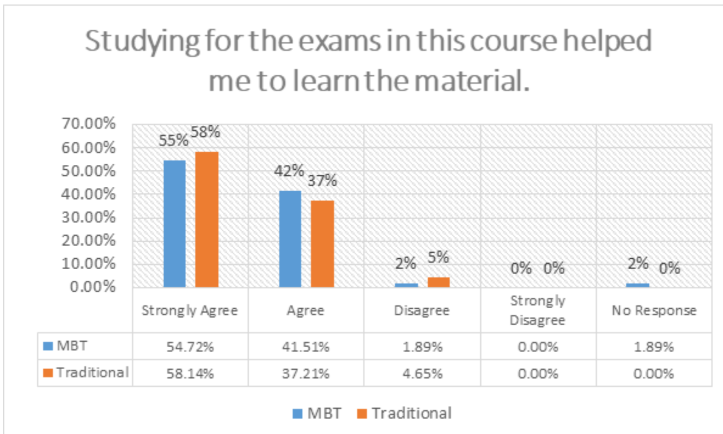
APPENDIX B

For each survey question analyzed, we provide a graphical output of survey responses as well as a summary table with the percentage for each response (with some responses combined).

Research Question 1: Does mastery-based testing impact how students approach studying?

- **Studying for exams in this course helped me to learn the material.**

Figure B1. Sample Responses to Survey Question 2



- **I relied mostly on memorizing solutions to earlier problems to prepare for in-class assessments.**

Figure B2. Sample Responses to Survey Question 7

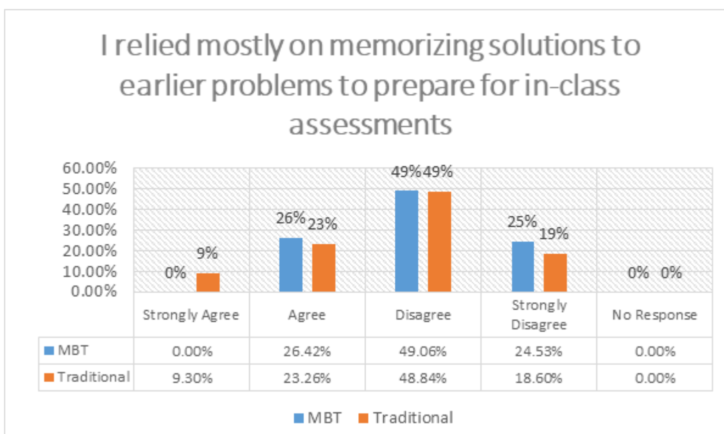


***Table B1.** Percentage of Survey Responses for MBT Group and Traditional Group*

	Agree/Strongly Agree	Disagree/Strongly Disagree	No Response
Studying helped learn course content	MBT: 96% Traditional: 95%	MBT: 2% Traditional: 5%	MBT: 2% Traditional: 0%
Relied on memorizing solutions to prepare	MBT: 26% Traditional: 33%	MBT: 74% Traditional: 67%	MBT: 0% Traditional: 0%

- **How many hours per week did you spend on this course outside of class time?**

Figure B3. Sample Responses to Survey Question 12

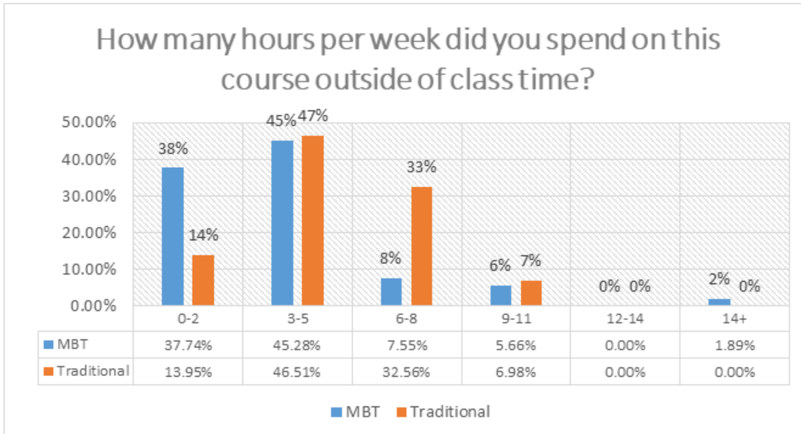
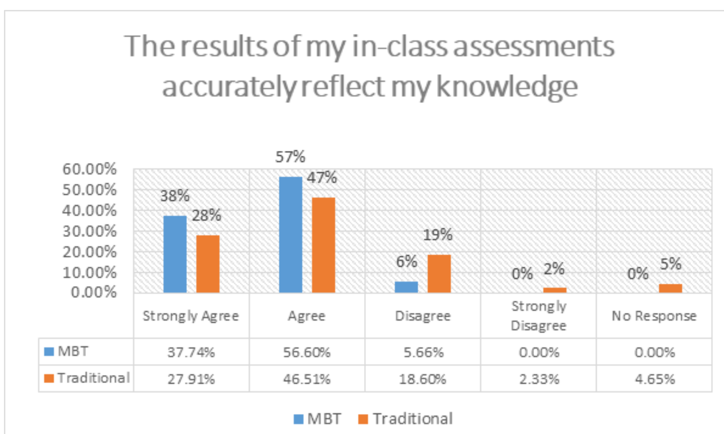


Table B2 . *Percentage of Time per Week Spent on the Course for MBT Group and Traditional Group*					
	0-2 hours	3-5 hours	6-8 hours	9+ hours	No Response
Mastery-based testing	37%	45%	8%	8%	2%
Traditional testing	14%	46%	33%	7%	0%

Research Question 2: Does mastery-based testing impact student reflection on their own knowledge?

- **The results of my in-class assessments accurately reflect my knowledge.**

Figure B4: Sample Responses to Survey Question 3



- **The in-class assessments deepened my understanding of the ideas in this course.**

Figure B5. Sample Responses to Survey Question 4

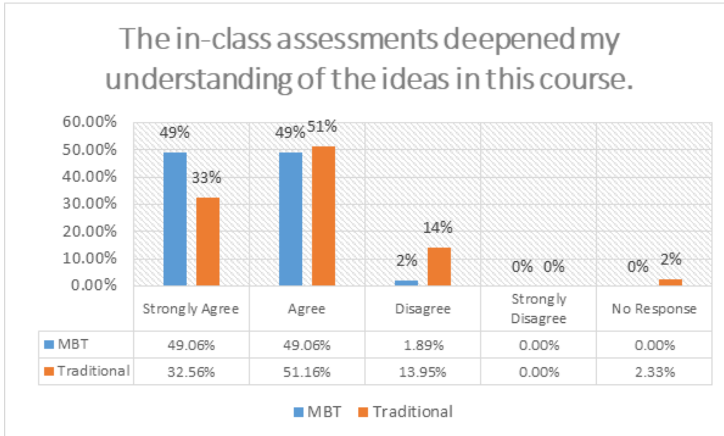


Table B3. *Percentage of Survey Responses for MBT Group and Traditional group*			
	Agree/Strongly Agree	Disagree/Strongly Disagree	No Response
Accurately reflect knowledge	MBT: 94% Traditional: 74%	MBT: 6% Traditional: 21%	MBT: 0% Traditional: 5%
Deepened understanding	MBT: 98% Traditional: 84%	MBT: 2% Traditional: 14%	MBT: 0% Traditional: 2%

Research Question 3: Does mastery-based testing impact end of semester content knowledge and grades?

Figure B6. End of Semester Grades for All Participants

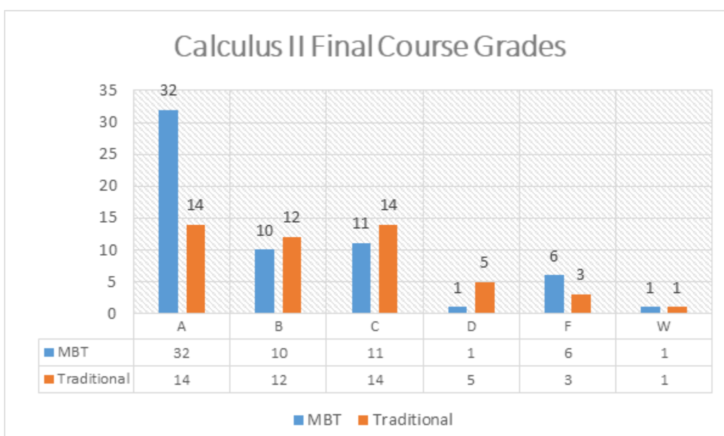


Table B4. *Percentage of Student End-of-Semester Grades for MBT Group and Traditional Group*					
	A	B	C	D	Not Passing
Mastery-based testing	53%	16%	18%	2%	11%
Traditional testing	29%	24%	29%	10%	8%