*Rebecca Swanson, COLORADO SCHOOL OF MINES, swanson@mines.edu*
*Aram Bingham, IIT-BOMBAY, aram@matmor.unam.mx*
*Megan Sanders, COLORADO SCHOOL OF MINES, sanders@mines.edu*
*Carter Moulton, COLORADO SCHOOL OF MINES, cartermoulton@mines.edu*

# Mastery-Based Testing in Linear Algebra: An Entry Point to Alternative Grading

## ABSTRACT

An increasing number of university faculty have become interested in alternative grading systems in recent years. While there is a growing body of research on such grading systems, implementation can be a challenge for instructors. Mastery-based testing is a grading system in which students learn from feedback and have multiple attempts to demonstrate proficiency in a written exam setting without penalty. Student exam scores are determined by the number of course learning outcomes successfully completed by the end of the course. Given that all other aspects of the course are left in place, this model can serve as a relatively simple entry point into alternative grading. We implemented mastery-based testing in three sections of a 16-week high-enrollment university linear algebra course and gathered data on student perceptions of and performance in this revised course. In this paper, we report on students' descriptions of the benefits and drawbacks of this system, their beliefs about final grades in this course compared to other courses, and their self-reported study habits.

## KEYWORDS

mastery learning, alternative grading, linear algebra, feedback loops, test anxiety

## INTRODUCTION

Over the past decade, there has been increasing discourse among university faculty and educational developers about the ineffectiveness of and potential harms caused by traditional grading practices (see Blum 2020; Clark and Talbert 2023; Feldman 2018; Nilson 2014; Schinske and Tanner 2014). By "traditional" grading practices, we mean systems where students earn points or percentages (including partial credit) and often have only one opportunity to demonstrate their understanding of course material through high-stakes exams worth a significant portion of their course grade (Clark and Talbert 2023, 11). This body of research emphasizes that traditional, point-based grading disincentivizes collaborative learning (Howitz, McKnelly, and Link 2021), stifles intrinsic motivation and creativity (Schinske and Tanner 2014), encourages shortcuts and academic dishonesty (Anderman and Murdock 2007; Blum 2020; Pulfrey, Buchs, and Butera 2011), does not provide meaningful, actionable feedback (Brookhart 2008), and constitutes a biased (Malouff and Thorsteinsson 2016; Moss-Racusin et al. 2012), inconsistent currency that ultimately flattens the nuances of student learning (Stommel 2023). Another major concern from our own teaching experiences is that awarding partial credit on high-stakes assessments is highly subjective and precludes an important learning opportunity for students. These issues led us to explore alternative approaches to grading important courses in the college mathematics curriculum.

Clark and Talbert (2023) define four pillars of effective alternative grading systems. We note that these alternative grading systems do not eliminate the final assignment of an A, B, C, D, or F grade but employ alternative assessment practices to determine that final grade. Following Clark and Talbert (2023), we will continue to use the term “alternative grading” to describe our work and that of others who do, in fact, use alternative practices in assigning grades to students without altering university-prescribed grading scales. The four pillars of effective alternative grading systems are: have clearly defined standards, provide helpful feedback, offer marks (brief indicators of progress) on individual assignments, and allow reattempts without penalty. These pillars work to construct a feedback loop for students and their instructors, in which students receive instructor commentary from which to learn before subsequent assessment attempts. Clark and Talbert give evidence of how these pillars provide students with a successful learning experience and address many of our concerns. We sought a grading system that would incorporate these four pillars while also scaling to our multi-section, high-enrollment mathematics courses.

Many of the systems we encountered, such as standards-based grading or ungrading, appeared daunting in complexity and instructor workload. In seeking a simple solution, we landed on mastery-based testing (Collins et al. 2019), a system where exam scores are determined by the number of learning outcomes a student successfully demonstrates at some point during the semester. In recent years, a small body of research has demonstrated the effectiveness of this approach (Collins et al. 2019; Harsy, Carlson, and Klamerus 2021; Harsy and Hoofnagle 2020). However, many other papers focus primarily on the implementation experience rather than the effectiveness (Halperin 2020; Harsy 2020; Linhart, Kulik, and Bangert-Drowns 2020; Mangum 2020), leading us to develop our own study.

Given that our engineering institution consists entirely of students in STEM fields, we wanted to know more about how the experiences of our students compared to those of students at other institutions reported in other studies of mastery-based testing. To investigate this, we conducted a study of 120 linear algebra students over a 16-week semester (see “Our approach” subsection for more information). The students in these courses were taught by one of the authors (Swanson), who is a member of the mathematics teaching faculty, and students were given both pre- and post-course surveys to assess the following research questions:

1. What were the benefits and drawbacks of this system for our students?
2. Did students believe their grade better reflected their learning when compared to other courses?
3. How does studying for this course compare to how students study for other courses?

We begin with a brief background on the history of mastery-based testing and relevant results. We follow with a description of our implementation and data collection methods before reporting the results of our study.

## BACKGROUND

### Mastery-based testing

The idea that student learning can be assessed by the demonstration of mastering explicit outcomes is not new. In the 1960s, mastery learning became more formalized, beginning with Bloom (1968). The premise of Bloom’s work was that, given enough time, all students can conceivably attain mastery of a learning task. In fact, he stated that “the grade of A as an index of mastery of a subject can, under appropriate conditions, be achieved by up to 95 percent of the students in a class.” Harsy and Hoofnagle (2020) give an excellent description of the history of mastery learning since Bloom’s work, focusing on a form of mastery learning course design which moves at a uniform pace but allows students to learn from feedback and demonstrate their learning at various points during the semester.

Multiple meta-analyses have demonstrated increases in student learning under mastery learning compared to control groups (Guskey and Gates 1986; Kulik, Kulik, and Bangert-Drowns 1990; Slavin 1987). Additionally, Guskey and Gates (1986) found that "students who learned under mastery conditions generally liked the subject they were studying more, were more confident of their abilities in that subject, felt the subject was more important, and accepted greater personal responsibility for their learning than students who learned under non-mastery conditions," supporting Bloom's (1968) predictions. Bloom also posited that a mastery learning system might have positive mental health effects. These results suggest the benefits of mastery learning systems go beyond mere exam performance.

The specific form of mastery grading that we utilized is called mastery-based testing (MBT) and is described in Collins et al. (2019). In this system, students have multiple attempts to demonstrate proficiency by individually addressing problems associated with a predetermined set of course outcomes. While the final course grade is still a weighted average of course components, such as homework, participation, and exams, the exam portion of the final grade is determined by the number of outcomes a student is able to "master" before the end of the semester. This does not mean that perfection is required, but rather a demonstrated proficiency with the use of a particular computational method, theorem, or concept (Clark and Talbert 2023, 9). There are numerous possible modifications to such a system as described in Collins et al. (2019), Halperin (2020), Harsy (2020), Linhart, Kulik, and Bangert-Drowns (2020), and Mangum (2020).

In previous studies, students reported that MBT has multiple benefits. Students feel that this system supports deeper learning and a more authentic experience in which mistakes are normalized and treated as learning opportunities. For example, Harsy and Hoofnagle (2020) found that students in MBT sections were more likely to believe that assessments deepen their understanding of ideas when compared to students in non-MBT sections. In Collins et al. (2019), students across multiple courses and institutions overwhelmingly reported that MBT helped them learn the material (82%) and prepared them for a variety of problems (84%). Students also feel that final grades better reflect learning in an MBT system (Collins et al. 2019; Harsy and Hoofnagle 2020). Given these benefits, Collins et al. suggest that the MBT approach may help students develop a growth mindset toward learning mathematics, particularly because students have many opportunities to overcome challenges. Even though Harsy (2020) found no statistically significant differences in mindset between clculus II students in MBT and non-MBT sections, a single intervention might not be enough to overturn years of conditioning.

Another student-reported benefit of MBT relates to decreases in test anxiety. Harsy, Carlson, and Klamerus (2021) found that 37% of students in non-MBT sections strongly agreed with the statement "I felt anxious before exams," as opposed to only 19% of students in MBT sections. Collins et al. (2019) also reported only 23% of students in MBT courses strongly agreed with the above statement. When asked to compare their experiences, 90% of students in MBT sections in Harsy, Carlson, and Kalmerus (2021) agreed that they feel less anxious before mastery-based exams compared to non-MBT exams.

While MBT may reduce student anxiety overall, Harsy, Carlson, and Klamerus (2021) also reported that some students in MBT classes said their anxiety levels increased in the middle of the semester. These may be students who either were unable to complete a sufficient number of outcomes early in the course or put off trying until later in the course. Bloom's prediction that mastery of a subject can be achieved by 95% of the students in a class did not account for the structural barrier that the semester length is fixed by an institution. The fact that students are taking other courses that may not offer the same flexibility could account for the challenge of procrastination. Harsy and Hoofnagle (2020) also list procrastination as a complaint found in some end-of-course student evaluations of teaching in MBT sections.

An MBT approach also seems to benefit student performance, though the research evidence measuring this impact is limited. In one study, a common final assessment was graded on a mastery scale; the average score was slightly higher in MBT sections compared to non-MBT sections, but this difference was not statistically significant (Harsy and Hoofnagle 2020). While many more students received As in MBT sections, a few more students also received Fs (Harsy, Carlson, and Klamerus 2021; Harsy and Hoofnagle 2020). Given that MBT is a form of mastery learning, one would expect the benefits to student performance would be comparable to those reported by Guskey and Gates (1986), Kulik, Kulik, and Bangert-Drowns (1990), and Slavin (1987). Yet the benefits of MBT for student performance cannot simply be explained by students spending more time engaging with the material—a concern raised by Slavin (1987). In previous work, most students in MBT sections reported spending only three to five hours a week studying (Collins et al. 2019) and reported spending less time studying than their peers in non-MBT sections (Harsy and Hoofnagle 2020). It may be that they spent less time studying because they either mastered outcomes quickly or benefited from clearer expectations. Despite this, students in MBT sections still struggle with the lack of partial credit (Harsy, Carlson, and Klamerus 2021). While partially knowing content is not the same as knowing it (Collins et al. 2019), students accustomed to partial credit in non-MBT mathematics classes are sometimes frustrated when their learning progress is not rewarded with some number of points.

### Our approach

Taken together, the above findings served as the foundation for our MBT approach. This work was done in the context of a public engineering institution in a suburban area in the United States (US). This institution has high research activity and an enrollment of approximately 5,700 undergraduates, most of whom complete their baccalaureate degree in a four-year period. The university offers a single upper-level linear algebra course that is required by several major programs, including computer science, electrical engineering, physics, and mathematics. In a typical 16-week semester, the institution offers eight or nine uncoordinated sections of linear algebra, each with 30–40 students. These sections are taught mostly by teaching professors—faculty members whose primary job duty is teaching. We implemented MBT in three of these sections, which were also taught in a partially flipped manner as described by Carney, Ormes, and Swanson (2015).

In our linear algebra implementation of MBT, students were given a list of 25 learning outcomes which corresponded to problem-types that could appear on an assessment. The students had between two and six opportunities to demonstrate mastery over each outcome in proctored, individual test settings spread over nine different assessment dates (see Figure 1). The instructor provided targeted feedback to the student for each attempt which did not meet the criteria for mastery. Once a student demonstrated proficiency on a given outcome, they no longer needed to attempt the outcome on future testing days. All assessment attempts took place during scheduled class meetings and were proctored by the instructor.

**Figure 1.** Detailed testing schedule

| | | | UNIT 1 | | | UNIT 2 | | | UNIT 3 | |
|---|---|---|---|---|---|---|---|---|---|---|
| WEEK | TIME PROVIDED | ASSESSMENT | Outcomes 1-3 | Outcomes 4-6 | Outcomes 7-9 | Outcomes 10-12 | Outcomes 13-14 | Outcomes 15-18 | Outcomes 19-22 | Outcomes 23-25 |
| WEEK 4 | 20 minutes | QUIZ 1 | ✓ | | | | | | | |
| WEEK 5 | 30 minutes | QUIZ 2 | ✓ | ✓ | | | | | | |
| WEEK 7 | 90 minutes | EXAM 1 | ✓ | ✓ | ✓ | | | | | |
| WEEK 9 | 20 minutes | QUIZ 3 | | | | ✓ | | | | |
| WEEK 11 | 30 minutes | QUIZ 4 | | | | ✓ | ✓ | | | |
| WEEK 12 | 90 minutes | EXAM 2 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| WEEK 14 | 25 minutes | QUIZ 5 | | | | | | | ✓ | |
| WEEK 15 | 50 minutes | RETEST DAY | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| FINAL EXAM PERIOD | 120 minutes | FINAL EXAM | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| TOTAL NUMBER OF ATTEMPTS GIVEN | | | 6 | 5 | 4 | 5 | 4 | 3 | 3 | 2 |

Image credit: Carter Moulton 2024

The portion of a student's grade (70%) determined by exam scores in previous editions of the course was replaced by a "mastery testing" grade, calculated as simply the number of outcomes for which the student demonstrated mastery divided by the total number of course outcomes. Regular written homework assignments were also assigned and traditionally graded, though each problem on a given homework was labeled with one of the course outcomes so students knew what to expect on MBT assessments. Each attempt at a learning outcome could receive one of three possible marks, presented to students as follows.

- **M: Mastery.** This is earned by completing the problem correctly and communicating work fluently by using proper terminology and notation.
- **MC: Communication Error.** This is earned by completing the problem correctly, but the work contains terminology or notation errors.
- **P: Progressing.** This is earned by submitting a problem that contains errors.

A student who was marked P on a given attempt was expected to reattempt that outcome on a later assessment. The MC mark made students eligible to submit a revision and reflection assignment which could convert their mark to M upon satisfactory completion, though no mastery credit would be granted until this condition was met. A "retest day" near the end of the term was provided so that students could reattempt any outcomes from the course ahead of the final exam date. The final exam was students' last opportunity to earn mastery on any outcome.

The details of our implementation choices adhere to the principles grounding the MBT literature, with one important nuance. Our MC mark plays a role similar to a category called "revisable" in other alternative grading schemes, though it is not always a part of MBT. We viewed the MC mark as a way to support student fluency with the novel language and notation of linear algebra while still rewarding those that demonstrated computational and conceptual proficiency. We believe it was an important part of the student experience, promoting productive engagement with course feedback loops.

When assigning marks to individual problems, the instructor's decisions were informed by the philosophy described in Collins et al. (2019) that portrays the withholding of mastery as a learning opportunity for the student. When deciding whether to ask a student to reattempt on a future assessment, a guiding question was, "will this student's understanding benefit from spending more time with this content?" Under this approach, we overlooked simple errors that were deemed irrelevant to the learning outcome being tested (e.g. arithmetic) if the student demonstrated clear conceptual understanding. Furthermore, we intentionally varied problems between assessments to promote a deeper level of content knowledge than rote memorization of past solutions could provide.

## RESEARCH DESIGN

We explored student perceptions and performance in our MBT course which ran from January to May of 2023 (spring 2023 semester). The research was approved as exempt by the university's Human Subjects Research committee. To gauge student perceptions, we conducted a pre- and post-course survey (see Appendix A), which included open-ended questions and rating scale questions that students answered on a 5-point Likert scale. For simplicity in reporting the responses to Likert scale questions, we will refer to both "strongly agree" and "agree" as simply "agree," and similarly for "disagree." Questions focused on student perceptions of the benefits and drawbacks of the MBT system, beliefs about grades, and anxiety around assessments (see Appendix A).

We administered the pre- and post-course surveys during class time, in the first week and the second-to-last week of the semester respectively. When administering the surveys, the instructor stepped out of the room, another member of the research team described the study to students and answered any questions, and students completed a consent form as the first page of the survey. A total of 112 students participated in the pre-course survey (93% of 121 enrolled students) and 108 participated in the post-course survey (90% of 120 enrolled students, after one withdrawal). Likert scale questions were summarized using the distribution of responses. Open-ended responses were coded using an approach similar to thematic analysis (Braun and Clark 2006). Two of the authors (Bingham and Swanson) read through the set of responses and identified initial themes. We then individually coded all the responses using these themes, meeting afterwards to compare our codes and to resolve any points of disagreement.

To assess the effect of our approach on performance, we compared student performance on specific exam questions from the sections of linear algebra taught in spring 2023 using MBT to sections of linear algebra that did not employ MBT but were taught by the same instructor (Swanson) in the previous semester.

## RESULTS AND DISCUSSION

### What were the benefits and drawbacks of this system for our students?

Overall, survey responses indicated that students were highly receptive to the MBT system, and we believe that the evidence shows that the perceived benefits of this new system significantly outweighed perceived drawbacks. We will present the findings on student answers by common themes, with the benefits and drawbacks in dialogue around particular features of our MBT implementation.

#### *Reduced test anxiety*

Reduced test anxiety appears to be a widespread experience among our students, with 79% of students agreeing with the statement "I feel anxious before exams in other courses," and 84% with the statement "I feel nervous about making mistakes on exams in other courses" on our post-course survey (see Figure 2). Reported test anxiety in other courses was largely consistent across the pre- and post-course surveys. In contrast, only 23% and 21% of students, respectively, agreed with corresponding statements when applied to the MBT linear algebra course on the post-class survey.

**Figure 2.** Reduced test anxiety in MBT vs. other courses

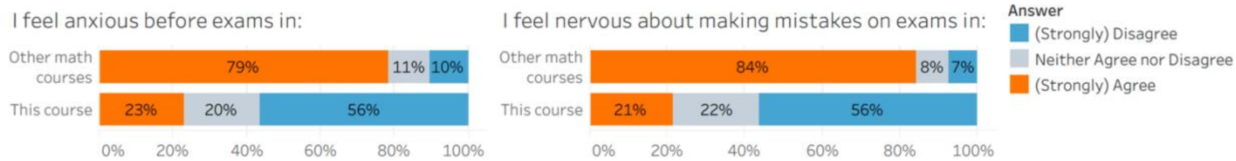


Image credit: Megan Sanders 2024

These survey results show that the system largely fulfilled one of its main promises, reduced test anxiety, as presented in Collins et al. (2019). In fact, our data show a significantly lower incidence of test anxiety for MBT compared to Collins et al. (2019; 23% vs 50% agreeing that they experienced test anxiety). Similar proportions of students in both studies reported feeling test anxiety in other, non-MBT courses (around 80%), confirming a significant improvement in this aspect of the student experience. We note that the finite number of opportunities to demonstrate mastery and the inevitable end date of the course do create pressure for some students, as can be seen in Figure 2. We did not ask how student anxiety levels changed over the course of the semester, but the persistence of some test anxiety in our course perhaps relates to the finding from Harsy, Carlson, and Klamerus (2021) that nearly 30% of students reported increasing anxiety levels as the semester progressed.

Nevertheless, 81% of students mentioned reduced stress and/or the incorporation of chances to fail in answering our open-ended question on the benefits of MBT. In this respect, we regard MBT as supporting a thesis of Bloom (1968) that mastery learning could provide mental health benefits to students because it offers them frequent opportunities for their self-development to be assessed and confirmed. As some reported: "[B]eing able to rectify previous mistakes and not be punished if you learn the content eventually is a really healthy way to go about learning which I admired." Additionally:

> The system really, really helped my testing anxiety and allowed me to just enjoy learning and not freak out if I wasn't able to get something right away. It motivated me to pursue feedback for my work and try to do better.

Overall, the data from our student surveys concord with our instructor perceptions of the student experience. By midway through the course, we sensed that most students believed they would be able to master an outcome within a few attempts and with concerted effort. Many who had been skeptics early on began to trust the system and engaged productively in the feedback loops.

*Procrastination*

The idea of multiple available assessment attempts for each outcome did lead some students to put off studying until the end of the semester. With any alternative grading system, instructors must be aware that their course competes for student attention. Students may not always responsibly manage the additional flexibility MBT offers when faced with inflexible deadlines from other courses. However, as with non-MBT courses, this can be managed to an extent by keeping track of which students have not mastered as many outcomes as their peers at various points of the semester and reaching out to them. Additionally, we noticed that many students did figure out how to work effectively in this system after a period of adjustment, exhibiting growth as self-directed learners. One student reflected, "It is also easier to slack off studying for this class overall, so I've learned a lot of self-control with my studying and doing it in a more timely manner than before."

For those that do not adjust to the testing format, a pile-up of unmastered outcomes at the end of a semester can become concerning. In our post-course survey, 16% of students mentioned this end-of-semester pressure as a drawback. Harsy and Hoofnagle (2020) did not report frequency data for this issue but acknowledged multiple students mentioning struggles with procrastination in post-course surveys of their MBT calculus II courses. While procrastination was a challenge for some of our students, it also simply took some of them longer to master outcomes than their peers, which could cause distress, especially as the course comes to an end: “If you don’t quickly get outcomes at the beginning of the course, it becomes really easy to feel overwhelmed and like most of your grade is in the air.”

As we do not have control over the length of the semester, we recognize that any deadline may cause anxiety for some students. This signals one of the practical limitations of mastery learning approaches within institutional contexts—though Bloom asserted 95% of students could eventually master any outcome, his hypothesis did not include any specified bounds on the time or effort necessary to reach this goal. The academic freedom instructors have in choosing our assessment practices may never entirely overcome certain structural barriers, but our data on student achievement (reported below) suggest that MBT gets closer to Bloom’s mark in one semester than non-MBT grading does.

#### *Assessment plan reflects the learning process*

One of the stated purposes of alternative grading systems like MBT is to better reflect learning as “a process, not a product” (Lovett et al. 2023, 3). Though we did not explicitly use the language of “growth mindset” with our students, the first day of the course included a discussion where we asked students to reflect on how they had learned a skill in the past and what the process was like (Talbert 2023). In the conversation, we emphasized responses that included some aspect of trial-and-error, coaching, feedback, reflection, and adjustment in response to failure. Over the course of the semester, we have observed that students look forward to the next opportunities for assessment, and they are eager for the feedback they receive, regardless of whether they earn mastery credit for a given outcome. In an open-ended question about the benefits of the system, 45% of students brought up the built-in chances for failure as something they appreciated, demonstrating their recognition of participation in the feedback cycle as a key part of their learning. One survey comment reads:

> The mastery-based grading system allowed me to learn the content of the course, fail horribly on most of the quizzes, learn from my mistakes, and correct my work on the exams. I definitely felt that I learned the concepts better in this format when compared to a traditional grading system.

In this respect, we believe that MBT delivers the promise of the mastery learning philosophy—most students are eventually able to demonstrate mastery on a large majority of the course outcomes and feel capable and empowered by their achievement at the end of the course. In the language of *Grading for Growth* (Clark and Talbert 2023), the system effectively incorporates the “reattempts without penalty” pillar common to many alternative grading systems. The radical normalization of making mistakes as part of the learning process is an important idea to impart upon students, and one which we hope they might carry across their academic careers.

A significant number of students (19%) also remarked unsolicited on the fact that the system emphasized learning over grades as a benefit. One student said: “I get to focus on the material itself rather than my grades. There has not been a point in the semester where I have thought about my grade which is very relieving." It is difficult to know to what extent student preoccupation with grades and points interferes with learning, but we found that with MBT, we experienced far fewer conversations with students arguing about exam scores as compared to when we taught the course previously without MBT. We believe centering course content and learning outcomes positively affects learning by encouraging students to productively engage with feedback loops.

As discussed in Joe Feldman's book *Grading for Equity* (2018), traditional grading is effective at sorting students according to who can learn course content quickly, but this may reflect and reinforce a host of inequities present in the student cohort when they enter the course, such as prior preparation and access to outside resources. We found a model of grading that provides room for making mistakes to be reflective of organic learning processes, and we were pleased to find the students recognized this benefit.

#### *Forgetting*

While some students worked to continually improve their understanding of a topic over the course of several attempts on a learning outcome, some experienced quick success. Given that students did not have to reattempt an outcome once it was mastered, these students were reasonably concerned about forgetting such material: on our open-ended question about drawbacks, 21% of students on the post-course survey mentioned concerns about the lack of reinforcement leading them to possibly forget earlier material or to not learn it to an adequate level. One student commented: "It could make the goal for mastering to get it once and forget it rather than to need that outcome again and again later in the course." However, this figure was down from 32% of students that had expressed this concern on the pre-course survey. While it is true that they only needed to prove their understanding for each outcome once in the course, in most courses, outcomes are assessed on exams only a small number of times—often only once or twice. For instance, in the prior semester of linear algebra (fall 2022), only 19 of the outcomes were assessed on any exam, and only four of them were assessed twice. Additionally, as linear algebra content builds upon itself, students had to demonstrate facility with early-semester material in order to succeed with later outcomes, leading several students to qualify such concerns, for example: "I expected that a potential drawback of this system would be lack of review of content that was covered early in the semester, but I didn't feel that this was a substantial problem, because of how interconnected everything was."

#### *Clear expectations*

Writing course learning outcomes (CLOs) is widely considered good pedagogical practice, and literature following Collins et al. (2019) mention clear course concepts as a characteristic feature of MBT. These articles did not investigate student perception of this feature, but our students appreciated the utility of clear learning outcomes. In response to our open-ended question on the benefits of the system from the post-course survey, 26% of students indicated that knowing how to focus their studying efforts on particular outcomes was helpful. We find the student testimonials the most poignant expression of this impact; for example, one student wrote, "I thought it relieved most [if] not all the stress that comes with exams. It also made it easier to study [because] the content was clearly outlined and defined." As another example:

> I always knew what I needed to study and what outcomes I knew in class. I found that exams were a lot less stressful and that it was easier to prepare for them. I loved how we had quizzes whose outcomes actually mattered.

The fact that students noted and appreciated the clear learning expectations synergizes with our other two main research questions. We further discuss the impacts these clear learning outcomes may have had on study habits in the final subsection below.

*Lack of partial credit*

The instructor communicated to students throughout the course that the demonstrated level of understanding required to earn mastery credit for a given learning outcome was high. Even so, the lack of points led some students to perceive their work on assessments as being closer to correctness than judged by the instructor. Some students may favor points earned through limited effort over the sustained effort required to achieve a high score and greater level of understanding. While only 10% of students expressed concern over the lack of partial credit in response to our open-ended question about drawbacks on the pre-course survey, that number jumped to 22% on the post-course survey. These comments mirror concerns brought up in prior articles, such as Harsy (2020), Mangum (2020), and Harsy, Carlson, and Klamerus (2021). None of these sources report the exact frequency of such responses, but Harsy, Carlson, and Klamerus (2021, 1084) mention concerns over lack of partial credit as their “most common complaint.”

Given that our students are accustomed to a system that awards partial credit, this concern is both expected and valid. However, as discussed by Collins et al. (2019), withholding partial credit is crucial to create learning opportunities that lead to comprehensive understanding. Although our students did not specifically mention “liking” the absence of partial credit, survey responses were still overwhelmingly positive about other aspects of the grading system. None of the comments describing lack of partial credit as a drawback indicated a clear preference for non-MBT grading, and several of these were qualified by statements that appreciated our intentions:

> I live on partial credit and the lack of it in this class was hard to get used to. With that said, because I had to get everything right, I feel like I had a better overall understanding of the content.

As we will explore below, student concern for a lack of partial credit is connected to a broader discussion about what grades mean. While our goal in this system is to connect grades to learning, many students believe effort should play a large role in their grades. As one student responded:

> Unless you show absolute understanding, any progress or better understanding than at the start is difficult to recognize in a final grade. I worry my grade in this class won’t reflect just how much time I’ve put into learning this and how much I’ve grown.

It is important to note that the post-course survey was administered before the last two assessment dates of the course and before grades were finalized. It is possible that these students’ partial credit concerns were alleviated by the time final grades were assigned.

*Grades align with learning*

As instructors, we view the lack of partial credit as a feature, rather than a bug. We were interested in testing the idea presented in Collins et al. (2019) that MBT incentivizes a deeper level of content knowledge, and that withholding credit for incomplete progress provides students with motivation to improve their understanding. While our survey did not directly ask students to compare this aspect of their experience to non-MBT courses, 84% of students agreed that the MBT system supported their learning of the material. Similarly, 75% of students agreed that the grading system deepened their understanding of the material. For comparison, 98% of respondents in Harsy and Hoofnagle's (2020) study agreed that MBT assessments deepened their understanding of course material.

Another source of evidence for increased achievement resulting from MBT comes from a comparison we made to two linear algebra sections from the semester prior (fall 2022, also taught by Swanson) that did not use MBT. The format of this course was identical, except with two traditional midterms and a final exam substituting for the mastery testing grade (70% of total). By mapping exam questions from this semester onto the learning outcomes used in the MBT version of the course, we could measure the proportion of students that demonstrated mastery over the outcomes at some point over the semester. We should note that only nineteen of the twenty-five learning outcomes appeared as comparable exam questions in fall 2022, and only four of these appeared more than once. Furthermore, the limited number of attempts and reduced number of outcomes given to the students in the previous non-MBT course, the lack of the list of clearly presented learning outcomes to orient their exam preparation, and the different time frames make this a weaker comparison than results given by Harsy and Hoofnagle (2020), who also found evidence of higher achievement among students in an MBT course. For 79% of the outcomes compared in our case (15/19), students performed better with MBT, meaning between 10% and 75% more students in the MBT version of the course demonstrated mastery at some point in the semester. See Appendix B for further details.

### Did students believe their grade better reflected their learning when compared to other courses?

Figure 3. Grades reflect learning in MBT more than other courses

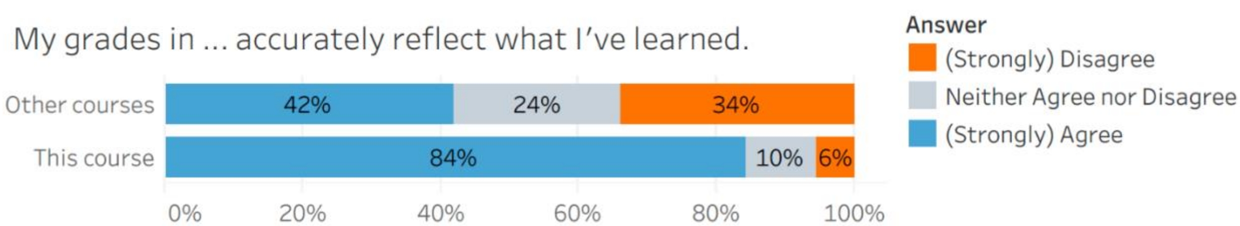


Image Credit: Megan Sanders 2024

Students' answers to multiple post-survey questions support the claim that our goal of better measuring their understanding of important course content was met. Around 42% of students agreed that their final grades in other courses accurately reflect what they learned, while that number rose to 84% for the MBT course (Figure 3). This is in line with the 81.3% that responded this way to a similar question in Collins et al. (2019) and not far from the 94% found in Harsy and Hoofnagle (2020). We note that the latter study did not include a neutral response as an option for this question, though our data still show a starker contrast with students' perception of non-MBT courses; 74% of students in a non-MBT graded calculus II course agreed that their grades reflected their learning in that study, versus only 42% of students in our study. Additionally, similar to Harsy, Carlson, and Klamerus (2021) and Harsy and Hoofnagle (2020), we found that the final grade distribution of our course showed more As (59% vs. 49% in the previous non-MBT semester), but unlike that previous work, we did not find a higher incidence of D/F/W grades compared to non-MBT semesters (6% MBT vs. 12% non-MBT); see Appendix C for further details.

This aspect of the student experience is closely related to the perceived fairness of the MBT system. Of the students we surveyed, 88% agreed that the system was fair. Our open-ended question, “What does your grade at the end of a course represent?” revealed similar evidence that students feel MBT gives a better reflection of their own learning than non-MBT courses. Our survey question did not specifically ask students to distinguish between “this course” and “courses in general,” but in our analysis we split the responses according to which context the student was describing—62 respondents (or 53%) for “this course” and 54 respondents (or 47%) for “courses in general.” Among those reflecting on linear algebra, 56% made comments indicating that their grade was a true measure of their understanding of the course content (Figure 4). This number dropped to 39% for those commenting on courses in general. Moreover, 20% of the group commenting on “courses in general” indicated that their grade was not in fact a good measure of their understanding, versus only 5% of those reflecting on linear algebra in the post-course survey.

**Figure 4.** Student perceptions of grades in MBT course compared to courses in general

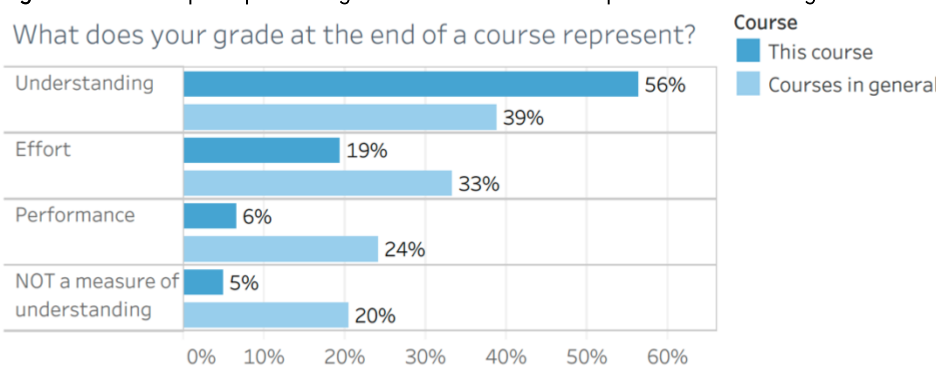


Image credit: Megan Sanders 2024

Related to our discussion of partial credit in the previous subsection, some students felt that their effort should be reflected in grades and the results of course assessments. As instructors, we believe that while effort is an important component of the learning process, grades for our courses should not be directly tied to effort in ways that deprioritize learning. Yet of those students who answered our survey question addressing “courses in general,” 33% indicated that their grades are reflections of invested effort in some way. This compares to only 19% that mentioned effort as a significant factor in their linear algebra grade. This doesn’t necessarily indicate that students expended diminished effort in linear algebra—some student comments indicated that in other courses, their grade was more based on the completion of tasks that took time and energy, but possibly lacked substantive learning outcomes: “[A grade] represents your constant effort in the class. However, I do not think that it should be like that. It should instead represent your mastery over the topic.” Other students contrasted effort invested in learning content and obtaining true understanding with the notion of unrepresentative performance on tests: “While I think [a grade] should represent what you learned, I think it represents how you performed on the final and midterms.”

Although 70% of each student’s final grade in our course was determined by their mastery testing grade, the multiple opportunities to demonstrate learning significantly mitigated such concerns. While around 25% of the students describing “courses in general” mentioned performance on exams as a dominant factor in their grades, only 6% of those discussing our course cited this as a key determinant of their Linear Algebra grade. On the other hand, 44% of students speaking of linear algebra described their grade as reflecting the number of course learning outcomes mastered, in language that contrasted with their test experiences from other courses. “The grade represents my success in this course, and my ACTUAL understanding, instead of the too common ‘hopefully put right answer on exam and move on.’” As another example:

> Typically, my grade at the end of a course represents what I knew on three days spread out over the semester. I think my grade in this class will actually represent what I have learned by the end of the semester.

Related to these responses, 18% of all students mentioned flexibility in when they are expected to learn a concept as a benefit of MBT, and 22% of respondents said that the system promoted depth in their learning in answer to our open-ended question on benefits. We find these results to be in concert with the differentiated test-taking experience referenced in the quotes above. Much of the student language describing traditional testing posits fleeting and superficial understanding that results from “cramming,” as opposed to the thorough, eventual understanding supported by MBT. We cannot claim this experience was universal, but we did find it to be common enough to warrant the continued utilization of this model.

### How does studying for this course compare to how students study for other classes?

We asked the above open-ended question directly of students in our post-course survey. After performing our thematic analysis of the responses, the most salient response by far was that students were better able to focus or target their studying (46% of respondents) than in non-MBT courses. As one student wrote: “While other courses may state outcomes, this system makes them matter more and it is clear why they exist.” By clearly communicating our learning outcomes, we also avoid some of the antagonism students have come to expect from their math instructors around what will show up on a test. While this dynamic is not solely a determinant of study habits, we consider it another benefit of MBT that students view instructors as being aligned with supporting their achievement:

> I definitely liked that all exam questions were directly related to both the outcome and past quiz or exam questions. Not having to worry about being tricked on exams made studying more straightforward and I was actually able to learn the content taught, not random small things that were never reviewed like I have had to in other classes.

Ironically, we have also heard this as an objection to systems like MBT from other instructors who favor cumulative, highly weighted (final) exams in non-MBT courses. These instructors’ intention is to induce students to conduct broad reviews over all the hidden nooks of course material. However, we believe that MBT does a better job of motivating learning habits like “spaced repetition” and “retrieval practice” (Clark and Talbert 2023, 34–35), which lead to lasting proficiency and understanding. Indeed, 12% of students responded that their studying was more spread out over the semester. Students are incentivized to invest significant effort in preparing for every assessment date, as the following quote illustrates.

> I definitely study much more intensely much more frequently for this class since every quiz carries just about the same weight as a test in my mind. So, it truly does help me

interact with the material more frequently on a deeper level which avoids the cramming that can come with the structure of a typical class.

Anecdotally, students' engagement with course content was more visible than in our non-MBT courses. Office hour attendance increased, and students came with clearer questions about what they needed to learn on the outcomes they were struggling with. Instead of talking about grades, we talked about linear algebra.

Nevertheless, survey responses varied on the overall amount that students reported studying as compared to other courses. For our open-ended question, 19% reported studying less than other courses, 11% reported studying more, and 16% said they studied about the same (as students were not asked to reflect directly on the amount of studying, these numbers do not add to 100%). We find no cause for concern in these responses, and our goal in utilizing MBT was not to make students work more or less than in a typical course. Instead, our findings seem to align with those of Harsy and Hoofnagle (2020) in allaying Slavin's (1987) concern that higher achievement in mastery learning systems stems primarily from a greater time investment. It is possible that upon encountering lower-stakes assessments, students reduced their invested study time, though this precise question has not been researched in an MBT context. We view the combination of clear expectations and helpful feedback provided by instructors as sufficient to overcome the additional work demanded by reattempts.

## IMPLICATIONS AND CONCLUSIONS

In this study we observed a variety of benefits for our students as well as notable shifts in what students reported about grades being accurate, what grades represent, and study habits—all of which are consistent with an enhanced student learning experience and greater overall learning achievement. While limited, perceived drawbacks of MBT reported by students shifted between the beginning and end of the semester: concerns about forgetting decreased while those related to partial credit and procrastination increased. We also note that the number of students reporting no drawbacks increased.

The focus of this study was related to the student experience, but the authors (Bingham and Swanson) also found the instructor experience in this and subsequent semesters to be a positive one. Given the clear connection between outcomes, assessments, and grades, we found conversations with students were focused on the learning outcomes as opposed to grades. Relatedly, we had a clearer picture of which concepts students were struggling with, as we could easily see which outcomes had lower numbers of "M" scores. The instructors also felt that the course grades assigned to students were a more accurate representation of their demonstrated learning. While many alternative grading practices work well with small class sizes, we were pleased to find that MBT was scalable to multiple sections. In fact, in a recent semester we have implemented MBT in seven sections of linear algebra with three distinct instructors coordinating grading efforts.

Finally, instructors in a classroom may be wary of trying something new if there is a chance it could negatively affect student evaluations of teaching. Harsy, Carlson, and Klamerus (2021) noted no significant difference between course evaluation and raw numerical scores for non-MBT vs. MBT groups, and our experience was comparable. In fact, one author (Swanson) found that in the semesters using MBT, almost every raw numerical score was about 0.2 to 0.3 points (on a five-point scale) higher than the averages from the previous nine sections in which she taught the course. Aside from a few technical challenges, mostly related to communicating grades to students within a Learning Management System, we have found the benefits for instructors strongly outweigh the drawbacks of MBT. We have found the overall time spent grading with MBT comparable to non-MBT courses—it is, however, "definitely time better spent" (Collins et al. 2019).

Given the growing body of evidence pointing to the benefits of alternative grading systems, we believe that the lack of widespread adoption is primarily due to a dearth of institutional resources that support instructors transitioning to grading systems like MBT. Indeed, an additional factor for our success was access to such resources. Support from the university's teaching and learning center for course development as well as the implementation of this study have been essential, as indicated by the authorship of this article. Furthermore, dedicated teaching faculty positions and a robust teaching postdoctoral fellowship program in the university's mathematics department have made investigations of innovative pedagogy more feasible. Training from external organizations like the Mathematical Association of America (MAA) and its professional development programs like Project/Section NExT were also helpful in orienting our study and navigating how to communicate about alternative grading systems with students.

While we have found positive results for MBT that align with the small body of previous work on this alternative grading system, there is still a need for additional studies, both at other types of institutions as well as in fields outside of mathematics, to determine the extent to which our results translate to other settings. Additionally, future work could explore some of the issues that arose for students in the semester we implemented MBT, such as procrastination and student concerns about partial credit. These issues are likely not unique to a MBT learning environment, but the way to address them may differ from non-MBT contexts. If MBT continues to exhibit positive benefits in future studies, this may also point to a need for additional resources and support for instructors wanting to implement MBT or other alternative grading models, as well as for departments and institutions to recognize faculty undertaking this work.

## AUTHOR BIOGRAPHIES

*Rebecca Swanson (USA) is a university distinguished teaching professor at the Colorado School of Mines where she teaches a variety of courses, implements flipped learning and alternative grading, and enjoys discussing student learning and participating in SoTL projects.*

*Aram Bingham (USA) is a mathematician and educator who was a teaching postdoctoral fellow at Colorado School of Mines from 2022–2024. Starting in 2025, he will be a postdoctoral fellow at the Universidad de Chile.*

*Megan Sanders (USA) is the director of the Trefny Innovative Instruction Center at Colorado School of Mines and earned a PhD in educational psychology from the Ohio State University.*

*Carter Moulton (USA) is a faculty developer at the Trefny Center at Colorado School of Mines. Carter holds a PhD from Northwestern University and previously served as a consultant and fellow at Northwestern's Searle Center for Advancing Learning & Teaching.*

## REFERENCES

Anderman, Eric, and Tamera Murdock. 2007. "The Psychology of Academic Cheating." In *Psychology of Academic Cheating*, edited by Eric Anderman and Tamera Murdock, 1–5. Elsevier Academic Press. https://doi.org/10.1016/B978-012372541-7/50002-4.

Bloom, Benjamin. 1968. "Learning for Mastery. Instruction and Curriculum." *Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints* 1 (2).

Blum, Susan, editor. 2020. *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead).* Morganstown, WV: West Virginia University Press.
Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. http://doi.org/10.1191/1478088706qp063oa.
Brookhart, Susan. 2008. *How to Give Effective Feedback to Your Students*. Alexandria, VA: Association for Supervision and Curriculum Development.
Carney, Deb, Nic Ormes, and Rebecca Swanson. 2015. "Partially Flipped Linear Algebra: A Team–Based Approach." *PRIMUS* 25 (8): 641–54. https://doi.org/10.1080/10511970.2015.1047545.
Clark, David, and Robert Talbert. 2023. *Grading for Growth*. Routledge. https://doi.org/10.4324/9781003445043.
Collins, J.B., Amanda Harsy, Jarod Hart, Katie Anne Haymaker, Alyssa Marie Hoofnagle, and Mike Kuyper Janssen. 2019. "Mastery-Based Testing in Undergraduate Mathematics Courses." *PRIMUS* 29 (5): 441–60. https://doi.org/10.1080/10511970.2018.1488317.
Feldman, Joe. 2018. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*. Corwin.
Guskey, Thomas, and Sally Gates. 1985. "A Synthesis of Research on Group-Based Mastery Learning Programs." Presentation at the *69th Annual Meeting of the American Educational Research Association* (Chicago, IL).
Halperin, Alexander. 2020. "Mastery-Based Testing in Calculus with a Final Exam Component." *PRIMUS* 30 (8–10): 1017–39. https://doi.org/10.1080/10511970.2019.1700575.
Harsy, Amanda. 2020. "Variations in Mastery-Based Testing." *PRIMUS* 30 (8–10): 849–68. https://doi.org/10.1080/10511970.2019.1709588.
Harsy, Amanda, and Alyssa Hoofnagle. 2020. "Comparing Mastery-Based Testing with Traditional Testing in Calculus II." *International Journal for the Scholarship of Teaching and Learning* 14 (2). https://doi.org/10.20429/ijsotl.2020.140210.
Harsy, Amanda, Christina Carlson, and Lauren Klamerus. 2021. "An Analysis of the Impact of Mastery-Based Testing in Mathematics Courses." *PRIMUS* 31 (10): 1071–88. https://doi.org/10.1080/10511970.2020.1809041.
Howitz, William, Kate J. McKnelly, and Renée D. Link. 2021. "Developing and Implementing a Specifications Grading System in an Organic Chemistry Laboratory Course." *Journal of Chemical Education* 98 (2021): 385–94. https://doi.org/10.1021/acs.jchemed.0c00450.
Kulik, Chen-Lin, James Kulik, and Robert L. Bangert-Drowns. 1990. "Effectiveness of Mastery Learning Programs: A Meta-Analysis." *Review of Educational Research* 60 (2): 265–99. https://doi.org/10.3102/00346543060002265.
Linhart, Jean Marie, James A. Kulik, and Robert L. Bangert-Drowns. 2020. "Mastery-Based Testing to Promote Learning: Experiences with Discrete Mathematics." *PRIMUS* 30 (8–10): 1087–09. https://doi.org/10.1080/10511970.2019.1695236.
Lovett, Marsha, Michael W. Bridges, Michele DiPietro, Susan A. Ambrose, and Marie K. Norman. 2023. *How Learning Works: Eight Research-Based Principles for Smart Teaching*. Indianapolis, IN: Jossey-Bass, An Imprint of Wiley.
Malouff, John, and Einar Thorsteinsson. 2016. "Bias in Grading: A Meta-Analysis of Experimental Research Findings." *Australian Journal of Education* 60 (3): 245–56. https://doi.org/10.1177/0004944116664618.
Mangum, Amanda. 2020. "Implementation of Mastery-Based Testing in Calculus." *Primus* 30 (8–10): 869–84. https://doi.org/10.1080/10511970.2019.1709587.
Moss-Racusin, Corinne, John Dovidio, Victoria Brescoll, Mark Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *PNAS* 109 (41): 16474–479. https://doi.org/10.1073/pnas.1211286109.
Nilson, Linda. 2014. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Routledge. https://doi.org/10.4324/9781003447061.
Pulfrey, Caroline, Céline Buchs, and Fabrizio Butera. 2011. "Why Grades Engender Performance-Avoidance Goals: The Mediating Role of Autonomous Motivation." *Journal of Educational Psychology* 103 (3): 683–700. https://doi.org/10.1037/a0023911.
Schinske, Jeffrey, and Kimberly Tanner. 2014. "Teaching More by Grading Less (or Differently)." *CBE–Life Sciences Education* 13 (2): 159–66. https://doi.org/10.1187/cbe.cbe-14-03-0054.

Slavin, Robert. 1987. "Mastery Learning Reconsidered." *Review of Educational Research* 57 (2): 175–213. https://doi.org/10.3102/00346543057002175.
Stommel, Jesse. 2023. *Undoing the Grade: Why We Grade, and How to Stop*. Denver, CO: Hybrid Pedagogy Inc.
Talbert, Robert. 2023. "A Growth-Focused Icebreaker." *Grading for Growth Blog*. https://gradingforgrowth.com/p/a-growth-focused-icebreaker.

# APPENDIX A

**Questions on the pre- and post-course surveys**

| | Pre | Post |
|---|---|---|
| What are/were the benefits of the mastery-based grading system used in this course? | X | X |
| What are/were the drawbacks of the mastery-based grading system used in this course? | X | X |
| What does your grade at the end of a course represent? | X | X |
| How does your studying for this course compare to how you study for other classes? | | X |
| I feel anxious before exams in other math courses. | X | X |
| I felt anxious before exams in this course. | | X |
| I feel nervous about making mistakes on exams in other math courses. | X | X |
| I felt nervous about making mistakes on exams in this course. | | X |
| My grades in other courses accurately reflect what I've learned. | X | X |
| My grades in this course accurately reflect what I've learned. | | X |
| This mastery-based grading system helped me learn the material. | | X |
| This mastery-based grading system deepened my understanding of the course content. | | X |
| I feel that this mastery-based grading system is fair. | | X |

APPENDIX B

## Mastery in non-MBT vs. MBT versions of linear algebra over the semester

| **Outcome** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Non-MBT %** | | 88.3 | 88.3 | | 58.4 | 62.3 | 88.3 | 66.2 | 57.1 | 41.6 | 63.6 | | 76.6 |
| **MBT %** | 100.0 | 98.3 | 98.3 | 98.3 | 95.8 | 92.5 | 100.0 | 80.0 | 99.2 | 92.5 | 98.3 | 97.5 | 100.0 |

| **Outcome** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Non-MBT %** | 20.8 | 80.5 | | 88.3 | 62.3 | 62.3 | 75.3 | 49.4 | | 63.6 | | 51.9 |
| **MBT %** | 95.8 | 76.7 | 98.3 | 74.2 | 91.7 | 75.0 | 91.7 | 35.8 | 97.5 | 77.5 | 78.3 | 26.7 |

Note: Percentages are the number of students demonstrating mastery over the outcome by the end of the course. Highlighting indicates outcomes that were assessed twice.

## APPENDIX C

### DFW rates for MBT course vs. non-MBT graded course

| | A | B | C | D | F | W | DFW |
|---|---|---|---|---|---|---|---|
| **Fall 2022 (Non-MBT)** | 49% | 23% | 16% | 6% | 1% | 5% | 12% |
| **Spring 2023 (MBT)** | 59% | 25% | 11% | 3% | 2% | 1% | 6% |