Exploring the Impact of Mastery Grading on Student Performance

Dr. Diana D Morris, University of Virginia

I am an assistant professor at the University of Virginia, and enjoy teaching a variety of math courses to undergraduates.

Dr. Hui Ma, University of Virginia

Dr. Hui Ma is an assistant professor of applied math at the University of Virginia. She holds a Ph.D. in applied mathematics from the University of Alabama at Birmingham. Her current research interest focuses on mathematics education and STEM education, in particular student-focused instruction, and learning-based grading.

Farzad Shafiei Dizaji, University of Virginia

Exploring the Impact of Mastery Grading on Student Performance – Work in Progress

Abstract: The introduction of mastery grading in Calculus I began in the fall of 2022 in our engineering school. We found that students in the mastery-graded course experienced reduced test anxiety, achieved higher letter grades through penalty-free reattempts, and felt more confident about their math ability. In the fall of 2023, we extended the implementation to additional sections and introduced several modifications to the grading system. The results on the common final exam showed a substantial benefit from mastery grading for students with the lowest diagnostic scores going into Calculus I. The final exam itself was structured to contain both straightforward, single-concept "Level 1" questions and more challenging, multi-step questions that blended multiple topics, "Level 2" questions. There was no significant difference in the performance between students with traditional grading vs. mastery grading for either level of questions.

Introduction

The vast majority of students entering the school of engineering at our university take Calculus II or III in their first semester. However, the number of Calculus I students rose sharply during the pandemic, ultimately reaching 23% in Fall 2023. Students who do take Calculus I enter with a wide range of high school math backgrounds and their struggles are reflected in a higher rate of D/F/W's (18% in Fall 2021) than students entering at other calculus levels.

Mastery grading was introduced in Calculus I in Fall 2022, largely to address disparities in the preparation of the students, and to combat anxiety and lack of confidence. Key features of mastery grading include breaking the course material into distinct learning outcomes. Students are allowed multiple attempts to demonstrate mastery in each learning outcome [1]. This approach aims to create a supportive and inclusive environment where students can achieve mastery at their own pace and foster a growth mindset by emphasizing continual learning over grades. Two sections were taught using the mastery grading approach, and two were taught using traditional grading. The mastery-graded group showed a significant decrease in test anxiety and felt more confident about their math ability [2]. Students appreciated the mastery grading approach for its benefits: 1) encouraging revisiting of ideas, promoting more practice, and enhancing knowledge retention; 2) providing stress relief through multiple reattempt opportunities, particularly beneficial in the first semester of college; and 3) offering helpful feedback through reassessments and office hours, supporting learning from mistakes, and improving understanding of the material. While there were measurable benefits for student's sense of well-being and confidence, no significant differences in performance on the final exam were observed.

In the fall of 2023, we extended the implementation to additional sections and introduced several modifications to the grading system. Three instructors taught five sections: two instructors taught three sections using the mastery grading approach. The third instructor taught two sections using traditional grading. To provide a more robust measure of student performance, a common mandatory final exam was given to all sections. This paper discusses the evolution of the grading

design and investigates the difference in students' performance in mastery-graded and traditionally graded courses by comparing final exam performance with scores from a diagnostic test in pre-calculus and Calculus I, given prior to the start of the semester. We also examined the effects of mastery grading on students' performance with different levels of questions on the final exam. As data becomes available, we hope to investigate how students from mastery-graded courses perform in the subsequent, traditionally taught, Calculus II course, compared to their peers in traditionally graded courses. This should offer insights into the enduring effect of the grading method on student success.

This study, which received approval from IRB, is currently ongoing, with data collection spanning the fall of 2023 and spring of 2024. The comparison of student performance will be based on their placement test scores and results from common exams. The examination of student performance will encompass both Calculus I in fall 2023 and the subsequent Calculus II course in spring 2024. The analysis of student's performance in Calculus II will take place following the conclusion of the spring semester.

Literature Review

In recent years, alternative grading approaches have gained significant traction in higher education. There is a wide range of practices in alternative grading systems, such as "Mastery-based Testing (MBT)" [3], "Standards-Based Grading (SBG)" [4], "Specifications Grading (Spec)" [5], or "ungrading" [6]. "Mastery grading" is used as an umbrella term for these alternative grading approaches. Mastery grading emphasizes students' demonstration of proficiency in specific learning targets rather than traditional letter grades based on a cumulative average of assignments and assessments. This shift encourages students to prioritize learning over grades, fostering continual improvement and growth throughout the semester. In their book "Grading for Growth" [7], David Clark and Robert Talbert comprehensively discussed the framework and benefits of alternative grading. They not only provided a guide to implementing alternative grading practices but also presented detailed case studies showcasing how instructors across diverse disciplines and institutions integrated alternative grading systems in various class formats.

Previous studies have highlighted several positive impacts, most notably a decrease in student stress. Instructors and researchers found that students experience less stress or anxiety during timed assessments, and they appreciate the opportunities to reattempt the concepts, without being penalized for early mistakes [8]. Additionally, some studies found that students enrolled in alternatively graded courses performed significantly better on the content assessment than those enrolled in traditional courses [9].

In Fall 2022, we also did a pilot study investigating the effectiveness of the mastery grading approach. We conducted pre-course and post-course surveys to assess students' beliefs about their math ability and math mindsets. The surveys included Likert questions based on a math self-efficacy scale [10] and math mindset questions (adapted from [11]). Math self-efficacy refers to an individual's confidence in their ability to succeed in mathematical tasks, categorized into mastery experience, vicarious experience, social persuasions, and physiological state (e.g., anxiety). The math mindset questions focused on whether students believed in fixed or growth

abilities in math. Additionally, the survey included questions about test anxiety, measuring the frequency of anxiety symptoms before, during, and after tests using a 5-item test anxiety inventory [12]. In our study, we found no significant differences in initial test anxiety levels between the traditional and mastery groups, with scores ranging from 5 to 20. However, by the end of the semester, the mastery group exhibited notably lower anxiety (9.78 vs. 12.64, $p < 0.001$). Within the mastery group, there was also a significant decrease in anxiety levels from an initial mean of 11.97 to 9.78 by the end of the semester ($p < 0.001$). Additionally, mastery students showed significant improvements in self-efficacy in mastery, vicarious experience, and social persuasion ($p = 0.005, 0.012, 0.018$), which was not observed in the traditional group. We compared students' placement scores between two groups and found no significant difference in preparedness (p-value=0.49). Despite the expectation that constant revisiting of topics in mastery graded sections would enhance retention and performance, there was no significant difference in performance at the end of the semester (p-value=0.86). However, the final grade distributions between the two groups indicated a considerable difference, with the majority of students in the mastery group obtaining an A. One of the major contributing factors is that students in the mastery-graded group had multiple opportunities to reattempt and demonstrate their understanding of the learning targets without any penalty before the end of the semester. Most students who are willing to put in the effort ended up achieving A- or A by the end of the semester. However, the final grades in the traditionally graded group were artificially lowered by the averaging process. The detailed results from our study in 2022 can be found here [2].

There remains a gap in longitudinal studies assessing the long-term impact of alternative grading. Specifically, we are interested in understanding how students who have been graded using alternative approaches perform in their subsequent advanced math and engineering courses.

Purpose and Research Questions:

To understand how mastery grading affects student performance, we looked into the following research questions:

1. How, if at all, does mastery grading affect student's performance across students' groups with varying levels of preparation? Which students benefit the most from the mastery grading approach?

2. How does the performance of students differ between groups on a common final exam containing questions with two levels of difficulty?

3. How do students from a mastery-graded course perform in a subsequent course, compared to peers in traditionally graded courses?

Implementation of Mastery Grading

The mastery grading scheme is structured around clearly defined learning targets, derived from Bloom's Taxonomy. For the Calculus I course in this study, we had 26 learning targets and 10 of which were designated as core learning targets. Class sessions involve pre-class preparation, collaborative problem-solving during class, and post-class online homework. Instead of

traditional exams, weekly checkpoints assess individual learning targets, each consisting of one problem (with multiple parts). Students receive marks of "mastered" or "progressing" based on their understanding. Students will have the opportunity to reattempt each learning target in the subsequent three checkpoints. A learning target is considered completed when students achieve the "mastered" mark twice. In addition to the weekly checkpoints held during class time, students also have the opportunity to demonstrate their mastery of a learning target during designated office hours. Final grades are determined based on the number of mastered learning targets, completed worksheets, and WebAssign performance, with the option to improve grades through a cumulative final exam. This approach aims to alleviate exam anxiety, promote deeper learning, and provide students with multiple opportunities to demonstrate understanding.

Obstacles and Challenges

As we implemented the mastery grading approach for the first time in Fall 2022, it's clear that we need to address some issues to make it work better for students (and for instructors too) and improve their learning experience. One significant issue observed was that due to the limited number of versions of each learning target problem, some students resorted to memorizing solution patterns through repeated assessments. Consequently, they achieved mastery grades without fully comprehending the underlying material. Additionally, many students procrastinated and only came for reassessments towards the end of the semester, leading to long queues during office hours. This not only burdened the instructor but also impeded the provision of individualized feedback, which was the intended purpose of office hour reassessments. The unlimited attempts policy failed to sufficiently motivate students to perform better initially and generated an excessive amount of grading. Furthermore, there are some practical and logistical issues we're facing, such as grading and keeping track of grades with tools like Gradescope, documenting reassessment grades, and ensuring transparent communication with students about their learning progress in a timely manner.

Modifications Made in Fall 2023

Reassessments play a vital role in the mastery grading approach, offering students opportunities to demonstrate their understanding after reflection and further study. However, it's essential to strike a balance between providing ample chances for improvement and maintaining accountability. Therefore, we reevaluated the number of reassessments allowed. By limiting the number of reattempts, we encourage students to put forth their best effort for each attempt, promoting a deeper engagement with the material and discouraging reliance on repeated attempts for memorization.

To streamline the reassessment process, each Learning Target appears on three consecutive checkpoints. Once students earn a Mastery mark twice on a target, it's considered completed. Rather than designated office hours, reassessment sessions are scheduled throughout the semester, each covering a list of learning targets. Additionally, two reassessment sessions are held at the end of the semester for final attempts. Students have five attempts for non-core targets and six for core ones. During reassessments, students simply sign in and specify which targets they want to attempt. TA office hours are also implemented to help students prepare and practice before reassessments. Furthermore, we created diverse problem versions for each learning target, coupled with limited attempts, to discourage memorization and promote genuine understanding of the material.

In Fall 2022, the final exam for mastery graded sections was optional, offering students a chance to improve their grades if they performed well. Without consequences for skipping or scoring poorly on the optional final exam, some students may not have felt motivated to give their best effort. Consequently, some submitted their final exams early upon realizing they were unlikely to achieve a high score. To address this issue and ensure a more accurate measure of student performance, we made the final exam mandatory in Fall 2023. Students receive a grade before the final based on three factors: the number of learning targets (including core ones) they have mastered, the completion of worksheets, and their overall average on WebAssign. The final exam now contributes to their course grade, potentially modifying it based on performance. For instance, scoring 88% on the final could raise a student's grade by one level, from a B to a B+. Conversely, scoring below 65% could lower their base grade by one level, from a B to a B-. See Table 1 for details.

Table 1: Course Grades, AFTER the Final

	Scores on the final needed to **move up one level**	Scores on the final needed to **keep your base grade**	Scores below this threshold will result in **moving down one level**
Before final: A-	≥ 85%	≥ 75%	75%
Before final: B+, B, B-	≥ 85%	≥ 65%	65%
Before final: C+, C, C-	≥ 80%	≥ 55%	55%
Before final: D	≥ 70%	≥ 45%	45%
Before final: F	≥ 65%		

Traditional Grading Scheme

Students in the traditional group were taught the same material as the mastery graded group. They had the same homework assignments and the same cumulative final exam. However, traditionally assessed students had three tests, all of which were graded with a traditional points-based and partial credit system. Their final grades were determined by the weighted average of worksheets, homework assignments, tests, and the final exam.

Methods

This study, which received approval from our Institutional Review Board, was carried out at the engineering school of a four-year, public university with roughly 22,000 students. To compare the effectiveness of the mastery grading system with the traditional grading system, we analyzed five Calculus I sections taught by three different instructors. Among these, two instructors taught three sections using the mastery grading approach, while the third instructor taught two sections using the traditional grading scheme. In fall 2023, 159 students enrolled in Calculus I. All 159 students were invited and 103 consented to participate in the study. Out of 103 students who consented, only 85 students agreed to the use of their artifacts and took the diagnostic test before the start of the semester. Among them, 29 were from the traditionally graded group and 56 were from the mastery graded group.

Data sources included students' placement test scores and final exam grades.

Results

Question 1: How, if at all, does mastery grading affect student's performance across students' groups with varying levels of preparation? Which students benefit the most from the mastery grading approach?

Most first-year students entering engineering school take a calculus placement test to determine which course in the calculus sequence is most appropriate. The test has three parts, covering pre-calculus, Calculus I, and Calculus II. The final placement decision is based on a combined consideration of the placement test, AP scores (if any), and student preference.

It is important to highlight that prior to comparing the performance of students at the end of the semester, we assessed whether there were differences in their preparedness based on placement scores between the two groups as a whole (mastery vs. traditional). The analysis revealed no significant difference in students' preparedness before the course started (p-value=0.32). Additionally, there were no significant differences found in students' performance at the end of the semester (p-value=0.76).

Students who entered in Calculus I are then categorized based on their placement scores into the following groups: 20 and under (Low), 21-25 (Medium), above 26 (High). A comparison of the placement test score and final exam score for each subgroup is shown in the following chart (Table 1). Only the scores for the pre-calculus and Calculus I portions of the placement test were considered, given the high proportion of Calculus I students who did not take Calculus II in high school. This included 45 questions with 30 pre-calculus and 15 calculus I questions, and each question is graded on a 1-point scale. Final exam score averages are provided for each range of placement test scores, along with the number of students in each category. See Table 2 and Figure 1 below for details.

Students in the lowest placement score category showed the biggest gains from the mastery grading system. Students at the high end of the placement tests showed the opposite result, performing better in a traditionally graded class. Students in the middle range showed very little difference.

Table 2: Students' performance on the final for different groups

Score on Placement Exam	**Low**	**Medium**	**High**
Average Final Exam (Traditional)	97.16 (8)	124.4 (13)	127.42 (8)
Average Final Exam (Mastery)	111.92 (23)	122.1 (17)	114.89 (16)

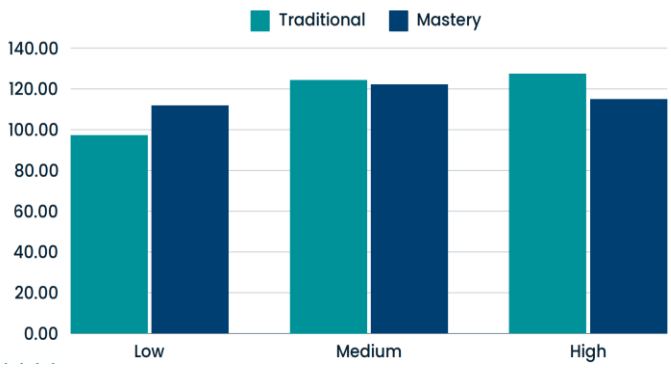

Figure 1

To address the possibility that students' pre-college preparation had more to do with their performance than the grading style itself, we created ordinary least squares regression models for each grading style, using placement test scores to predict final exam scores. One point of interest for us is whether the mastery grading approach can liberate student performance from the effects of the high school experience, i.e. whether students can flourish independent of background in this grading environment. The results of the linear regression were not statistically significant for the mastery graded sections, with a p-value of 0.173. In other words, students' knowledge prior to entering college was not useful in predicting their final exam results for this group. This was surprising, since we expected student pre-knowledge to be highly influential.

For the traditionally graded sections, we found a p-value of 0.044, suggesting there is reasonable evidence that the model is predictive. In other words, for this group, better placement scores (understood as better high school preparation) are associated with better performance on the final exam. The output of the regression is shown in Figures 4 and 5 below. The coefficient 2.14 for the independent variable (the placement test) indicates that for each point increase in the placement exam score, the model predicts a 2.14-point increase in the final exam score.

Given the size small of the data set and the very low R^2-value (0.14 for the traditional group), we are cautious about drawing conclusions, but in future work we will continue to explore whether mastery grading is associated with less connection between students' placement scores entering college and first semester final exam scores.

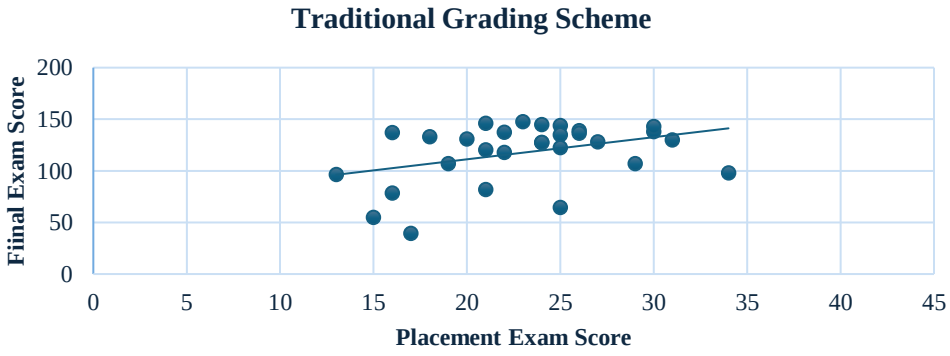

Figure 2. Placement exam score vs final exam score for traditional group

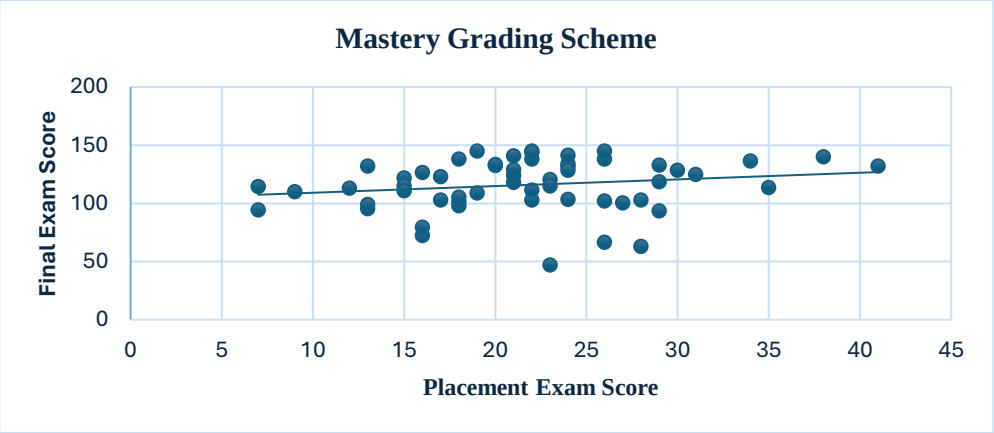

Figure 3. Placement exam score vs final exam score for mastery group

Regression Statistics	
Multiple R	0.37722428
R Square	**0.14229816**
Adjusted R Square	0.11053143
Standard Error	27.4355017
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3371.72794	3371.728	4.479471	**0.043666243**
Residual	27	20323.08241	752.7068		
Total	28	23694.81034			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	68.2935579	23.91230309	2.856001	0.008153	19.22956474	117.3576
Placement	2.14347806	1.012758009	2.116476	0.043666	0.065470267	4.221486

Figure 4. Regression Output for Traditionally Graded Group

Regression Statistics	
Multiple R	0.1846684
R Square	**0.0341024**
Adjusted R Square	0.0162154
Standard Error	21.836207
Observations	56

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	909.0801291	909.0801	1.906548	**0.17303559**
Residual	54	25748.27701	476.8199		
Total	55	26657.35714			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	103.3915	9.487840307	10.89726	2.99E-15	84.3695279	122.4135
Placement	0.5736038	0.415420733	1.380778	0.173036	-0.2592646	1.406472

Figure 5. Regression Output for Mastery Graded Group

Question 2: Is there a difference in performance between the two groups in the different types of questions on the final exam?

The final exam was divided into two groups of questions. There were thirteen Level 1 questions (single-concept and straightforward) and five Level 2 questions, which required multiple steps and crossed topics. Because the mastery grading approach limited the scope of questions on each assessment, we were unsure how these students would perform on problems requiring synthesis of several strategies. No meaningful differences were observed between the two groups in either type of question. The chart below (Table 3) shows the average percentage score in each category of question in each group.

Table 3: Level 1 and Level 2 Question Performance

	Level 1 Questions	**Level 2 Questions**
Traditional Grading	80.71	76.98
Mastery Grading	79.84	75.16

Question 3: How do students from a mastery-graded course perform in a subsequent course, compared to peers in traditionally graded courses?

We want to see how well students who took mastery-graded courses do in the next Calculus II class compared to students who took traditionally graded courses. This will help us understand how the grading method affects student success in the long run. We will analyze students' performance in Calculus II after the spring semester ends.

Discussion

In this study, our primary focus was on assessing the impact of mastery grading on student performance across various levels of preparedness and identifying which students benefited most from this approach. Despite making the final exam mandatory this year to obtain a more robust measure of student performance, we have not yet observed significant differences in their overall performance.

However, the data does indicate a notable advantage from mastery grading for students with the lowest diagnostic scores entering Calculus I. They outperformed their counterparts in the traditionally graded group within the same category. These students entering the course felt less prepared, leading to lower confidence and increased anxiety about their grades. They required significant support. Mastery grading likely provided the necessary support by allowing them to learn at a slower pace initially and providing penalty-free reattempts. This is an encouraging result as it suggests that mastery grading might address the needs of students who are less prepared. By offering a supportive environment and opportunities for penalty-free reattempts, mastery grading helps these students to overcome initial challenges and achieve better understanding and performance.

It's also worth noting that we conducted the same course surveys as last year. Pre-course and post-course surveys included Likert questions on their beliefs about their math ability using a math self-efficacy scale [10], and math mindsets (modified from [11]). Our findings from a repeated survey showed similar outcomes to the previous year, as we discussed in the literature review section. Students in the mastery-graded group experienced reduced test anxiety, achieved higher letter grades through penalty-free reattempts, and felt more confident about their math ability. These outcomes suggest potential benefits for their overall well-being and confidence.

Due to the nature of the assessments of mastery grading approach, which focuses on assessing specific skills or concepts in isolation, there was uncertainty regarding the performance of students when faced with problems that necessitated the synthesis of multiple strategies. We plan to incorporate additional course components, such as applied problems or projects, alongside assessments of individual learning targets for future iterations of the course.

Moreover, we found that the current structure of the course may inadvertently lead to a lack of challenge for high-performing students. By expanding the scope of course components to include more diverse and challenging tasks and assessments, we are hoping to make sure that all students, regardless of their level of mastery, are sufficiently engaged and challenged throughout the course.

Limitations and future work

The Calculus I course in this study was taught by three different instructors. Due to logistical constraints, it wasn't possible to have the same instructor teach both the mastery and traditionally graded sections concurrently. As a result, various factors could influence students' performance and their overall learning experience within each group. Future work will involve completing the analysis and investigating how students' mastery grading experience affect their performance and experience in the subsequent calculus II courses. We also hope to track, in future cohorts of students, whether the observation that placement tests better predicted final exam scores for students within the traditionally graded sections than those within mastery graded sections, holds on a larger scale. This would open the possibility that mastery grading "unlinks" the effects of pre-college preparation to college success. This next phase of investigation becomes particularly significant as we await the availability of additional data. By extending our inquiry beyond the immediate impact, we hope to gain insights into the longer-term effects and implications of the mastery grading approach on students' progress.

Reference

[1] Campbell, R., Clark, D. and OShaughnessy, J., 2020. Introduction to the special issue on implementing mastery grading in the undergraduate mathematics classroom. *Primus*, *30*(8-10), pp.837-848.

[2] Ma, H., 2023, June. Mastery Grading Approach in a Calculus Course. In *2023 ASEE Annual Conference & Exposition*.

[3] Harsy, A., 2020. Variations in mastery-based testing. Primus, 30(8-10), pp.849-868.

[4] Selbach-Allen, M.E., Greenwald, S.J., Ksir, A.E. and Thomley, J.E., 2020. Raising the bar with standards-based grading. Primus, 30(8-10), pp.1110-1126.

[5] Carlisle, S., 2020. Simple specifications grading. Primus, 30(8-10), pp.926-951.

[6] Kohn, A. and Blum, S.D., 2020. *Ungrading: Why rating students undermines learning (and what to do instead)*. West Virginia University Press.

[7] Clark, D. and Talbert, R., 2023. Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education. Taylor & Francis.

[8] Lewis, D., 2022. Impacts of Standards-Based Grading on Students' Mindset and Test Anxiety. Journal of the Scholarship of Teaching and Learning, 22(2), pp.67-77.

[9] Katzman, S.D., Hurst-Kennedy, J., Barrera, A., Talley, J., Javazon, E., Diaz, M. and Anzovino, M.E., 2021. The effect of specifications grading on students' learning and attitudes in an undergraduate-level cell biology course. *Journal of Microbiology & Biology Education*, *22*(3), pp.e00200-21.

[10] Usher, E.L. and Pajares, F., 2009. Sources of self-efficacy in mathematics: A validation study. *Contemporary educational psychology*, *34*(1), pp.89-101.

[11] Dweck, C.S., 2006. *Mindset: The new psychology of success*. Random house.

[12] Taylor, J. and Deane, F.P., 2002. Development of a short form of the Test Anxiety Inventory (TAI). *The Journal of general psychology*, *129*(2), pp.127-136.