Advances in Engineering Education

2024: VOLUME 12 ISSUE 2

Evaluation of Self-Assessment "Ungrading" Practices in a STEM Course

MINERVA BONILLA
Department of Construction Science
Texas A&M University

DANIEL FINDLEY
Institute for Transportation Research and Education
Department of Civil, Construction, and Environmental Engineering
North Carolina State University

ABSTRACT

Educators and institutions have considered and continue to explore alternatives to measure students' learning and track their performance. An alternative that started to gain popularity due to its effectiveness in promoting learning engagement, equity, and inclusion, and helping mitigate concerns due to mental health was "ungrading." Ungrading is a pedagogical approach that does not necessarily eliminate grading, instead, it promotes frequent feedback and engagement with students on their work to ensure the course learning goals have been met. Unfortunately, studies measuring the effectiveness of this method targeting STEM majors with a significantly large number of students enrolled have not been published. Therefore, this study evaluates how students' performances are impacted by the incorporation of "ungrading" in a civil engineering course offered to undergraduate students. This study utilized an "ungrading" practice, known as self-assessment, for the midterm examination and students' participation was optional. To track students' performance, and the effectiveness of the method, the student's overall course performance, and student feedback were evaluated to see if there was any difference in grading between those students opting to participate compared to the ones who did not. Findings indicated that grade performances and course engagement improve by participating in "ungrading."

Key words: Ungrading; Students Performance; Diversity, Equity, and Inclusion; STEM Courses

BACKGROUND

In 1785 institutions decided to introduce a grading scale system to measure students' performance in a more structured method (Espinola 2018; Gorichanaz 2022). However, when grading systems

were introduced, these could not aid students' development of intellectual character (Schneider and Hutt 2014). Researchers like Elbow (1993) and Kohn (1993 and 2018) argue that our current grading system focuses on including ranking, sorting, rewarding, and punishing students which discourages students from challenging themselves intellectually (Milton et al. 1986; Pulfrey et al. 2011, Schinske & Tanner 2014; Supiano 2019; ATL 2020; Gorichanaz 2022).

Other researchers consider grades to be problematic for not incentivizing learning, are vague, often inconsistent, provide little valuable feedback, and inaccurately assess students' learning (Brennan & Magness 2019; Blum 2020; ATL 2020; Stommel 2021). In a traditional grading system, instructors provide students with formative rather than summative feedback (CITLS 2022). These traditional grading system practices are considered damaging to students' mental health (Eyler 2018; ATL 2020; Supiano 2022).

Studies performed by Bloodgood et al. (2009) and the American Institute of Stress (2019) reported that 8 out of 10 college students in the United States report stress that negatively impacts their sleep and health, which affects their ability to learn. This stress can lead to mental health issues such as anxiety, depression, substance abuse, and suicide (Pascoe et al. 2020; Aslanian & Roth 2021). In addition, other studies such as the one published by Bouchrika (2020) have suggested that the traditional university grading system can be a source of stress and anxiety for students, which can negatively affect their physical and mental health."

Traditional grading systems also promote an unhealthy learning environment between students and instructors. In a study performed by Ginexi (2003), it was found that there is a correlation between students' grades and the course evaluations they provide to the instructor. This research suggests that students often use evaluations as a means of revenge. This is because students most often attribute their lack of success to the instructor's ineffective teaching (Guberman 2021).

On the other hand, the concept of "ungrading" as a pedagogical approach has been proven to create a more effective learning environment that enables instructors to focus more on supporting learning (Wettergreen et al. 2018; ATL 2020; Stommel 2020). This is done by eliminating grades from the student evaluation process and instead, providing meaningful qualitative feedback that promotes a learning environment where students can take risks, fail, and improve upon their work (ATL 2020; Jarvis 2020).

Ungrading also allows students to develop intellectual character and help them build the skills that will serve them to succeed in their careers (Gorichanaz 2022). Doing so allows students to receive formative feedback that promotes meaningful reflections that encourage learning beyond the boundaries of completing a set of tasks on an assignment (Stančić 2021; CITLS 2022; Koehler & Meech 2022). From the mental health perspective, research on alternative approaches to student evaluation has shown that "ungrading" practices enhance students learning and improve health and well-being (Spring et al. 2011; Gorichanaz 2022).

The concept of "ungrading" as a pedagogical approach started to gain popularity in 2015 (Jarvis 2022). However, this was not widely accepted by all universities or all departments (Burke 2020; Jarvis

2022). In 2020, due to the COVID-19 global pandemic, institutions started to re-evaluate and reflect on the concept of grades which has allowed the opportunity for some instructors to experiment with alternative pedagogical approaches such as institution-wide pass/fail to grade (Burke 2020; Veletsianos and Houlden 2020; Goldrick-Rab 2021; Ashby-King 2021) or the incorporation of "ungrading" practices at some liberal arts colleges and institutions such as Brown University (Gorichanaz 2022).

Hypothesis

The purpose of the study is to evaluate students' performance based on a pedagogical approach called "ungrading." Previous researchers have argued that 'ungrading' practices can provide a more humane and inclusive learning experience for students, as they shift the focus from grades to learning and foster a more collaborative and supportive classroom environment (Sharp 1997; Greenberg et al. 2022) and focus on student learning and growth. On recent research done on Science, Technology, Engineering, and Mathematics (STEM) courses, it was found that ungrading is an alternative assessment approach that seek to shift the focus from grades to learning and feedback in education. Ungrading has been adopted and experimented with by some instructors in various STEM disciplines, such as chemistry, mathematics, and physics, with different methods and outcomes (Jarvis 2020; Talbert 2022; Open Education 2020). These sources show that ungrading can have positive effects on students' learning and well-being in STEM education, such as fostering deeper learning, reducing stress and anxiety, enhancing metacognition, self-evaluation, and growth mindset, and increasing motivation, engagement, and agency. Jarvis (2020), Talbert (2022), and Open Education (2020) also acknowledge some of the challenges and limitations of ungrading in STEM education, such as aligning with professional practice, meeting institutional expectations, communicating with stakeholders, and ensuring quality and rigor. Therefore, it is suggested that ungrading is not a one-size-fits-all solution, but rather a flexible and adaptable strategy that requires careful planning, implementation, and evaluation in different STEM contexts.

As previously noticed, there has been little information on the incorporation of ungrading on STEM courses. At the time of this publication, there has been no published research, small article, or preliminary context available on the incorporation of ungrading on STEM courses with large numbers of student enrollment. Jarvis (2020) said that almost any time a talk about incorporating "ungrading" for STEM courses is brought up by him, he gets a comment suggesting this could not work. The perception that STEM courses require a more traditional and clear structure to evaluate students' performance is strong in STEM majors. However, studies have shown that traditional grading practices induce a loss of learning motivation, induce students to compete more, and decrease the relevance of instructor feedback (Greenberg et al. 2022). This is because students focus more on earning a grade than learning.

Multiple studies have identified benefits that prove "ungrading" offers more advantages than disadvantages. A summary of these findings is displayed in Table 1. In this table, we see the advantages and disadvantages of "ungrading" for students and the instructor's points of view on the impact on academic, personal, and social life. These studies show that "ungrading" promotes learning by nurturing practices such as work collaboration and promoting engagement between students and their instructor over feedback provided (ATL 2020; Guberman 2021; CITLS 2022; Greenberg et al. 2022). Lastly, even though ungrading presents some disadvantages, these might be eliminated or reduced if "ungrading" practices are properly delineated and executed.

Table 1. Advantages and Disadvantages of "Ungrading".

		Student	Instructor	Source
Advantages	Academic	• New learning habits. • Promote creative work. • Encourage a growth mindset.	• Promotes a forward-looking approach rather than a retrospective one. • Simulates a more real-world (i.e. business, industry) learning environment. • Students produce higher-quality work.	ATL 2020; CITLS 2022; Greenberg et al. 2022
	Personal	• Reduces mental health burden by reducing stress related to grades. • Promote student well-being and creative thinking. • Improve students' recognition of the value of their learning.	Focus on providing feedback rather than penalizing students for missing points.	ATL 2020; CITLS 2022; Jarvis 2020
	Social	• Promotes better communication. • Reduce competitiveness. • Improves classroom climate. • Aid to minimize inequities arising from socioeconomic, racial, gender, and other social forms of difference.	• Promotes better communication. • Promote a culture that encourages learning from struggle and failure, rather than punishing it.	ATL 2020; CITLS 2022; Greenberg et al. 2022; Guberman 2021
Disadvantage	Academic	• Students with a better understanding of the course material often assess their work more negatively. • Women often under-grade themselves relative to men.	• Ungrading practices are not always supported by the university policies • Ungrading is not a perfect approach to student evaluation. It must be coupled with questions and considerations that address effective learning environments.	ATL 2020; Jarvis 2020
	Personal	Ungrading massively reduces student anxiety, but students must trust the instructor. If students feel that instructors are only pretending to let them assess themselves, then ungrading will not work.	• Might require additional work to change the structure of the course. • Additional work to providing feedback and reviewing students' work.	Jarvis 2020
	Social	None identified by available research.	None identified by available research.	N/A

Types of "Ungrading" Approaches

The success of an "ungrading" approach is linked to proper execution. Multiple "ungrading" approach methods can be implemented in a course. Therefore, the term "ungrading" means different things to different people (Flaherty 2019; Jarvis 2020; Greenberg et al. 2022). Learning to identify the difference between these methods is essential to select the method that fits best with the class goals and objectives. In an article published at the Academy for Teaching and Learning at Baylor University (ATL 2020) six methods that can incorporate "ungrading" practices were identified. These methods are:

1. Evaluative, qualitative feedback
2. Contract grading
3. Self-evaluation
4. Portfolios
5. Student designed rubrics
6. Collaborative assessments

The first method *"Evaluative, Qualitative Feedback"* proposes the elimination of letter grades. Instead, the assessment process should focus on assessing the learning goals (i.e. goals related to reading, writing, discussion, research, and projects) for the student. In this method, the instructors should provide constructive feedback that offers new opportunities for the students to improve their work on the subject area. This is a method typically ideal for subjective courses and in combination with other "ungrading" methods.

"Contract Grading" allows the instructor to work as a facilitator and mentor. Instructors provide feedback to help students in their learning and allow for more creative work rather than penalize them. The idea behind *Contract Grading* is that instructors provide a form of evaluation that allows students to "contract" for a particular grade in a course. For example, if a student wants an 'A' in a course, the student must complete more assignments than a student contracting for a 'B' or 'C.' By doing so, this method reduces stress due to grades. This method is ideal for writing assignments and other subjective courses.

The third method, *"Self-Assessment,"* focuses on getting students involved in the assessment process by having them participate in the evaluation of their work. In this method, the instructor helps students assess the learning process and improve/acquire knowledge by self-evaluating their work. This process allows students to practice metacognition which gives them an awareness of their thought processes and understanding. To get the most out of this method, Stommel (2021) proposed that instructors should allow the students the freedom and autonomy to grow and learn, by just providing constructive feedback and critical reflection to guide the students in the process of evaluating their work based on self-growth. This method is ideal for courses and can also be combined with other "ungrading" methods.

In the method *"Portfolios,"* the students and instructors revisit and assess the entire learning experience throughout the course. Students are responsible to collect, annotate, and grade their

assignments, projects, and exams. The focus of the "portfolio should be on evaluating students' growth throughout their comprehension, application, evaluation, and learning success. To assess the work, the instructor should provide a questionnaire that would help the students reflect on the quality of their work. There was no indication of what type of courses could benefit more from this method.

The fifth method *"Student Designed Rubrics"* involves students in the process of constructing grading rubrics. By doing so, students can take ownership of the grading process and their work Jarvis (2020). The last method *"Collaborative Assessments"* proposes a collaborative assessment. This method is recommended for group activities or assignments, class quizzes, exams, presentations, or writing projects. The idea is that students should work in groups to share ideas, aid themselves in their learning process, and benefit from one another. This method promotes collaboration among students and increases conceptual understanding, retention, problem solving, and critical thinking skills (Gilley & Clarkston 2014).

The literature reviewed to perform this study evaluated multiple journals and publications that utilized various "ungrading" approaches. A summary of the methods adopted in this study is presented below in Table 2. This table provides the identification of the college where the course was taught, the level of the students (graduate or undergraduate), and the length of the course. In addition, it provides a small description of the "ungrading" method utilized, the goal of the study, and the findings. A total of three out of the seven studies performed "ungrading" in STEM courses. The findings indicated that students are more reluctant to the "ungrading" approach in comparison to a traditional grading system. At the same time, "ungrading" allows students to reduce their anxiety towards grading and encourages students to engage deeply with the course material by nurturing learning and course engagement.

Based on the findings from the literature, this research identifies the lack of assessment of the "ungrading" approach on large STEM courses due to the perception that this approach is not structured to properly assess students' learning. Therefore, this work proposes the evaluation of the "ungrading" approach for a large undergraduate level course with an enrollment of 72 students. The students enrolled in this course are in the junior and senior levels in the Department of Civil, Construction, and Environmental Engineering at North Carolina State University.

METHOD

Before the study began collecting data, approval from the North Carolina State University (NCSU) Institutional Review Board (IRB) was requested to perform human subject research. The approval was granted to perform research based on the evaluation of students in the CE 305 Traffic Engineering

Table 2. "Ungrading" Methods Adopted in Studies.

Course Description	Method	Goal	Results/Findings	Source
Students enrolled in a 16-week semester Michigan State University College of Education - Graduate level course.	Collaborative assessment and students designed rubric for all course assignments.	Provide the instructor's perspective on implementing "ungrading" practices during COVID-19.	Ungrading enabled classes to adapt and to better extend the care that teachers and students needed at the time of remote learning due to COVID-19.	Jennings 2021
Students enrolled in a 16-week semester Purdue University College of Education - Graduate level course.	Evaluative, qualitative feedback, self-evaluation, and collaborative assessment.	Investigated how graduate students navigated participation in self-assessment in an online course.	Advanced students struggled with the responsibility of establishing appropriate participation goals and managing their efforts to gain the most from their participation. Therefore, instructors must support learners to navigate an ungraded approach to course participation.	Koehler and Meech 2022
Students enrolled in a 16-week semester Drexel University College of Computing & Informatics - Undergraduate level courses.	Reflection-based self-evaluation in all assignments in the course.	Student experiences with "ungrading" course structure.	Students considered the system to be more difficult to navigate in comparison to a traditional grading system, making it difficult to see where they stood or how they were doing in the course. However, this method provided time to reflect on their work, deepen their learning, to enrich the communication that fosters a learning community within the classroom.	Gorichanaz 2022
Students enrolled in a 5-week summer session University of Tennessee and Carson-Newman University College of Psychology - Graduate level course.	Collaborative assessment in all assignments in the course.	Evaluated "ungrading" practices at the graduate level	After overcoming disorientation on the new pedagogical approach, students considered that "ungrading" provides value in collaboration, nourishes creativity, and learning, and allows interactive feedback from peers and instructors.	Greenberg et al. 2022
Students enrolled in a 16-week semester Keene State College of Biology and the University of Mary Washington College of Digital Studies - Undergraduate level courses.	Self-evaluation for the midterm exam and course assignment.	Evaluated "ungrading" practices and their effectiveness in online and STEM courses.	Ungrading reduces students' anxiety and nurtures learning and course engagement and promotes equity in the learning process.	Jarvis 2020
Students enrolled in a 16-week semester Rose-Hulman Institute of Technology College of Biomedical Engineering and Humanities and Social Sciences - Undergraduate level courses.	Evaluative, qualitative feedback, and collaborative assessment of assignments, quizzes, and exams in the course.	Evaluate the effect of "ungrading" practices concerning student learning and engagement with course material.	Ungrading allows for deep learning and encourages students to engage deeply with course material. At the same time, it discourages students from doing the bare minimum to "pass" the course without a clear indication that they have understood the content.	Dosmar and Williams 2022
Students enrolled in a 5-week summer session Purdue University College of History - Undergraduate level courses.	Self-evaluation and contract grading in all assignments in the course. This course did not have exams or quizzes.	Identification of the challenges facing the expansion of "ungrading" practices: the lack of traditional evidence of effectiveness.	Students felt confused by the structure of "ungrading" practices however, they saw the value in their learning process and the effectiveness of "ungrading" practices.	Guberman 2021

course. CE 305 is an undergraduate level course offered to students in the Civil, Construction, and Environmental Engineering Department at North Carolina State University. CE 305 is a three-credit hour undergraduate level course with a grading structure composed of four homework assignments (55%), a midterm (20%), and a final exam (25%). This study was performed during the 16-week long, Fall 2022 academic term with a total enrollment of 72 students. A total of 68 (54 menand 14 women) signed a consent form and agreed to participate in this study.

During the first class of the semester, the instructor explained class requirements and grading structure, and the concept of "ungrading" was introduced (based on processes proposed by Koehler & Meech 2022). It was explained that "ungrading" would only be used for midterm examinations and participation was optional. The instructor then explained the importance of "ungrading," and the rules, and procedures that they were expected to follow. This information was also repeated during a class period before the midterm exam and on the date, the instructor returned the exams with feedback.

Ungrading Procedures

To perform this study, we applied the "Self-Assessment" method, which allows the students to get involved in the grading process and evaluate their work. Other than the "self-assessment" activity on the mid-term exam, the course was delivered in a similar manner to other semesters. On the midterm exam, students were asked to do their best on every problem to fully answer each question and show their process for solving the problem. Each question on the exam had a confidence level (see Figure 1) represented by emoji. This idea was inspired by the studies performed by (Jarvis 2020 and Blum 2020) and was used with the purpose of helping students evaluate their work and indicate their confidence in solving the problem. Students were asked to place a checkmark or "x" on the line corresponding to the emoji that best represented their confidence in responding to the question. All students taking the midterm completed this section.

After the exam was completed, the exam was reviewed by the instructor and returned to students with some hints, feedback, or input on each question that was incorrect, which is a practice also utilized by (Guberman 2021). A table similar to the one in Table 3 was provided to the students

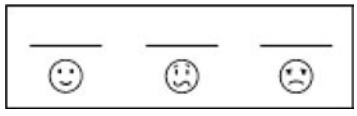


Figure 1. Confidence Level Used on Midterm Exam.

Table 3. Example of "Ungrading" Form for Midterm Exam [Provided on the Exam].

Exam Question (Points)	Student	Instructor	+/– Points	Final Score
1 (5)				
2 (5)				
3 (10)				
4 (20)				

on the last page of the exam, which is meant to be left blank during the exam and only completed after the instructor has returned the exams with feedback. This table was included for the purpose of facilitating grading. The first column had the exam questions and in parenthesis was the total maximum points a question was valued to and the standard deviation of the grades the students received based on the instructor's score (the standard deviations were not provided during this process because it was not computed at that point in the process). The second column was for students to assign the points they think they earned on each question. The third column had the instructor's assessment. The fourth column was to record any additional or reduced points (on questions with errors) and the final column had the final score students earned per question. This final score was based on an average of the student and instructor scores and +/- points. Part of the student/teacher agreement was that if a student's assessment differs by ≥ +1 standard deviation, one point will be subtracted from the final score for each standard deviation above the instructor's score. When the student's assessment differs by < 1 standard deviation, one point will be added to the final score. The overall timeline included: 1) students completed the exam, 2) the instructor provided written feedback and documented the assigned score (but did not communicate the score to the student) in the next class after the exam, 3) students reviewed their errors and provided a score (first column in Table 3), 4) students returned the exam to the instructor within three classes of having the exam, 5) instructor completes the remaining three columns and returns the exams to the students.

The students received their exams with feedback only and they were given a total of three class periods (approximately 1.5 weeks) to review their work. During this time, the students were asked to assign the points they think they earned on each question and justify their answers. After the review period was over, if a student would not like to participate in "ungrading" they would simply keep their exam. The students who opted to assess their work were able to justify their answer, if needed, on the last page of their exam with the grade they assigned themselves in the "Student" column in Table 3. After the exam was returned to the instructor, the instructor added the information in the remaining three columns. A sample of how this grading process was completed is shown in Table 4.

Table 4. Example of Completed "Ungrading" Form for Midterm Exam.

Exam Question (Points, St Dev.)	Student	Instructor	+/– Points	Final Score
1 (5, 1.5)	5	5	0	5
2 (5, 1.1)	5	5	0	5
3 (10, 2.4)	6	8	+1	8
4 (20, 2.3)	18	12	–2	13

Data Evaluation

Data collection and analysis started after midterm grades results were completed. The midterm performance analysis was first evaluated with the purpose of identifying any trends in students' records. Even though "ungrading" practices were used only for the midterm exam, the research team wanted to know if partial incorporation of "ungrading" practices can have an impact on student's performance. Prior to the final examination and the end of the course, students were provided with the analysis of their performance in the midterm exam, and a consent form was distributed requesting their permission to use their CE 305 grades for this study. We then waited until the semester concluded to review the consent forms and construct an analysis of the overall course performance.

In addition to evaluating students' performance on the midterm exam, and overall course performance, this study evaluated students' feedback. At the end of the semester, students are asked to anonymously provide feedback about the course structure, teaching performance, and overall course experience. We utilized this feedback to learn about students' experiences (good or bad) in relation to participating in the self-assessment activity on the midterm. Our goal was to see if utilizing alternative pedagogical approaches such as "ungrading" enhance learning and promote course material retention. The following section provides the findings of this study based on midterm performance and overall class performance.

RESULTS

The results presented in this section evaluated the student's performance of the 68 students who signed the consent forms. Their performance was evaluated based on their midterm exam grades and the student's performance on other class assignments, final exams, and the overall course was evaluated.

Midterm Performance

After the review period was over, 8 students decided not to participate in the "ungrading" review process utilizing the "self-assessment" method. A total of 59 students participated in self-assessment, and the following results were obtained from this sample.

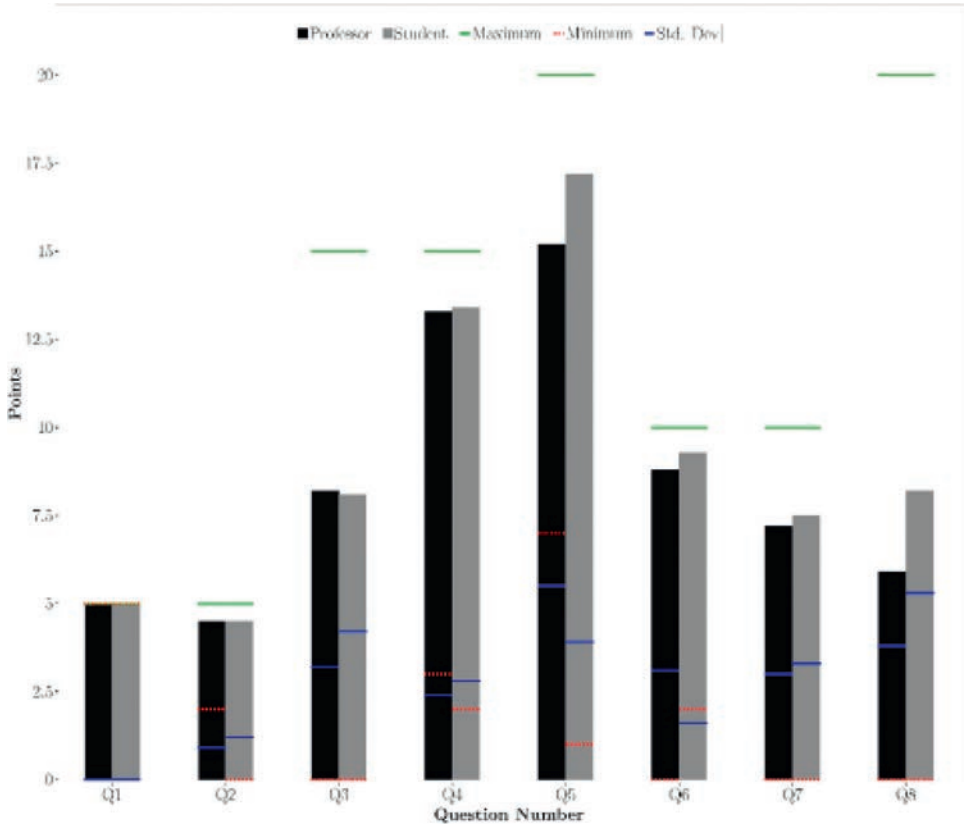


Figure 2. Midterm Grade Distribution per Exam Question.

The first analysis was the evaluation of grade distribution per exam question. These results are displayed as the average and standard deviation (SD) of the points earned by the students on each of the 8 questions in the exam. Figure 2 displays the average and SD of the grades that the instructor assigned and it also displays the grade distribution assigned by the students. In questions such as Q1 and Q2 students graded themselves similar to the grade assigned by the instructor. The only question where the average of the student's grades was lower than the instructor's scores is Q3 and in Q4, the students graded themselves slightly higher than the instructor. It can also be observed that students give themselves higher marks on questions Q5, Q6, Q7, and Q8.

The results of the student's confidence level per question on the midterm exam are displayed in Figure 3. In general, the results show some correlation between the assignment of grades and confidence level. It can be observed that on questions Q1, Q2, and Q4 where the students' average is similar to the instructor, the student's confidence level is high for questions Q1, and Q2 but relatively moderate for Q4.

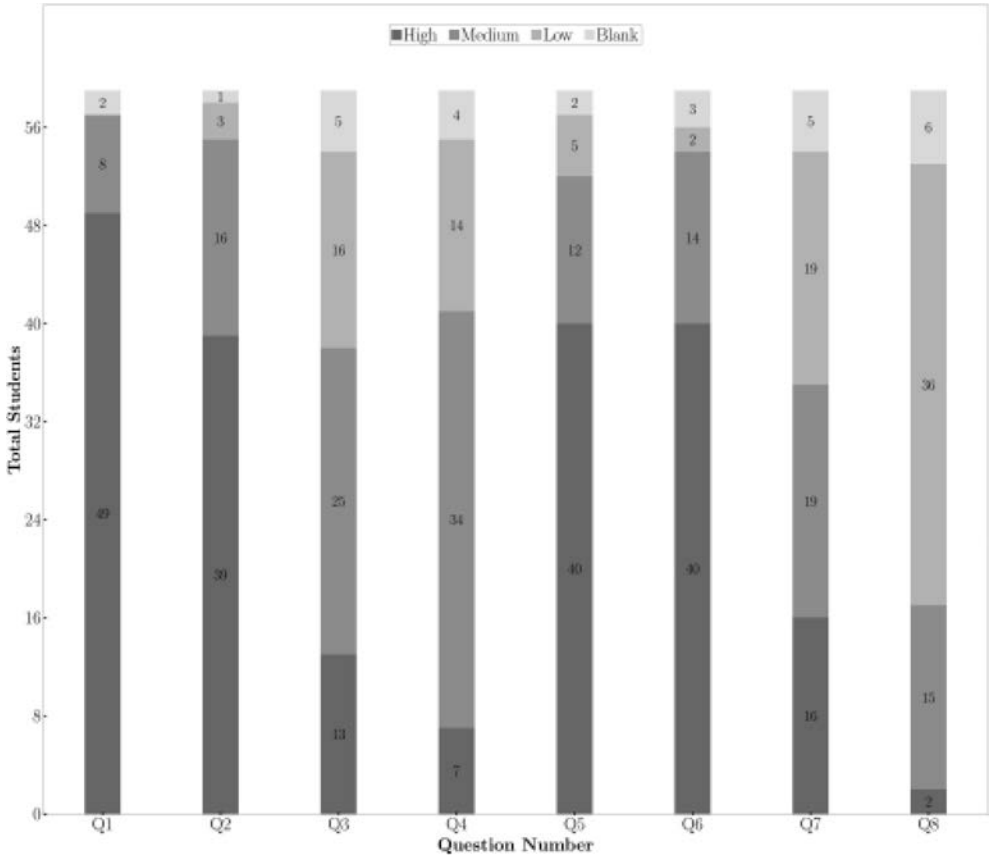


Figure 3. Students' Confidence Level per Question on Midterm Exam.

In Q3 (See Figure 2) the average grade given by the instructor was 8.2 and the average given by the students was 8.1. The confidence level for Q3 where students graded themselves lower than the instructor is mainly in the medium to low spectrum. On average, the student-assigned score was related to their confidence in solving the problem. Lastly, on the questions where the students graded themselves higher, we see a mixture of results. For questions Q5 and Q6, the student's confidence level was high, however, the grade difference between the instructor to student grade was more substantial. In Q5, the instructor assigned an average of 15.2 whereas students assigned themselves 17.2. For Q6, the instructor assigned an average of 8.2 whereas students assigned themselves 9.3. For question Q7 where we observed an overall medium confidence level, the students still graded themselves higher than the instructor but the grade difference was little. In Q7, the instructor assigned an average of 7.2 whereas students assigned themselves 7.5.

On the contrary, the question with the lowest confidence level was Q8. From these results, it can be observed that the students' grading approach shifted substantially. In general, the students

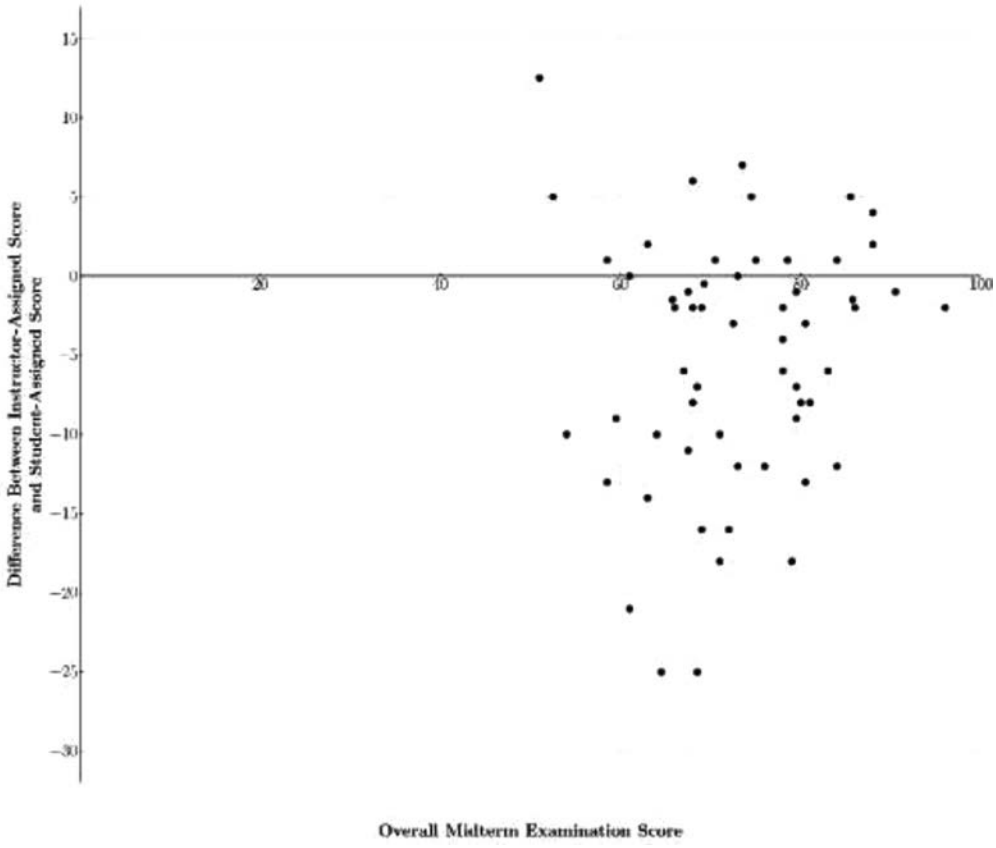


Figure 4. Record of Students' Score Change after Participating in the Self-Assessment.

assigned themselves higher scores on a question where they had the least confidence (and the lowest instructor-assigned scores, as well, though the actual instructor-assigned score was unknown to the students during their review of the exam). This discrepancy could be attributed to the student's lack of understanding of how to solve the problem. At the time students participated in the self-assessment they were provided with the correct answer, but not the fully worked-out solution. That was part of their assignment, to evaluate their work, see where they went wrong

Figure 4 displays the differences in the student-assigned and instructor-assigned scores (y-axis) and the final score (x-axis) on the midterm examination. Overall, there was an average 5-point difference between the student-assigned and instructor-assigned scores. Lower overall midterm examination scores had larger differences in the instructor-assigned and student-assigned scores. Exams with higher scores had fewer errors and therefore, fewer opportunities for scoring and differences in scoring.

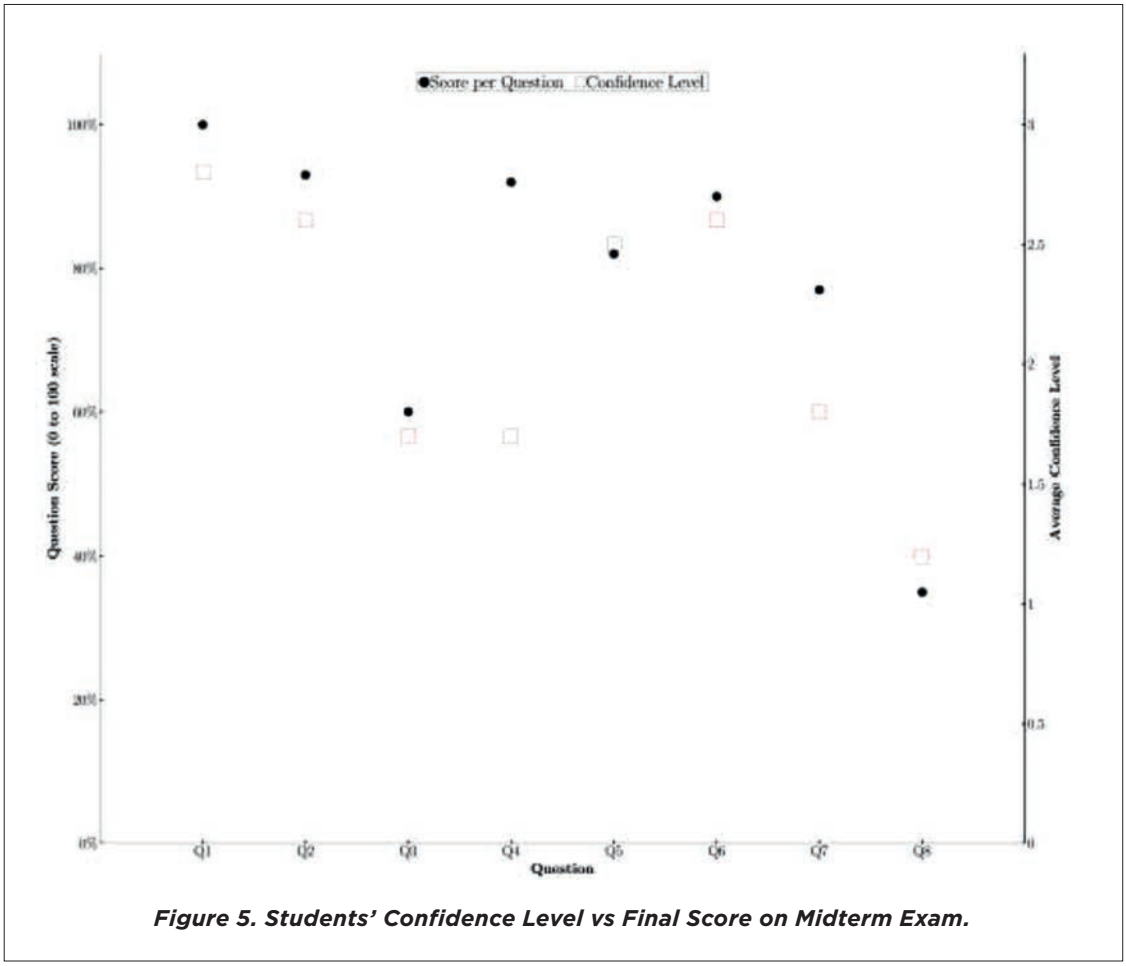


Figure 5. Students' Confidence Level vs Final Score on Midterm Exam.

As previously explained, students' grade distribution was not consistent with the confidence level that they provided on each of the midterm questions. Therefore, another important metric to consider was the evaluation of student's confidence levels concerning the final score they earned. This evaluation was made with the purpose of better understanding if the confidence level is linked to students' performance. Figure 5 shows the correlation between both metrics.

The results indicate that when a student's confidence level is relatively high, the scores earned per question are also high. When the student's confidence level is lower (i.e. Q3 and Q8) the score per question is also relatively low. These patterns are intuitive and expected, likely representing well-written questions and well-prepared students.

However, there were also two data points with divergent results in terms of confidence and scores. These were Q4 and Q7 where we observed that students' confidence level is lower and the score per question is higher than expected relative to the pattern from the other questions. In the instructor's opinion, these questions, in particular, focused on requiring students to apply critical

thinking to answer these questions and were not provided with a similar example of these questions during class (i.e., there were no clear comparisons in their notes or textbook to refer to). The students, on average, performed higher than their confidence level on both problems, which may indicate that they were able to use reasoning, logic, and their understanding of the material to solve the problem, but they did not have a commensurate level of confidence that they had successfully solved the problem.

Student's Performance on Homework Assignments

A total of four homework assignments were given to the students. These assignments were related to the material learned in class and the overall class performance was favorable. In Figure 6 we can see the results from these metrics for those students who participated in "ungrading" and those who did not. In general, the average score for students who participated in "ungrading" is higher.

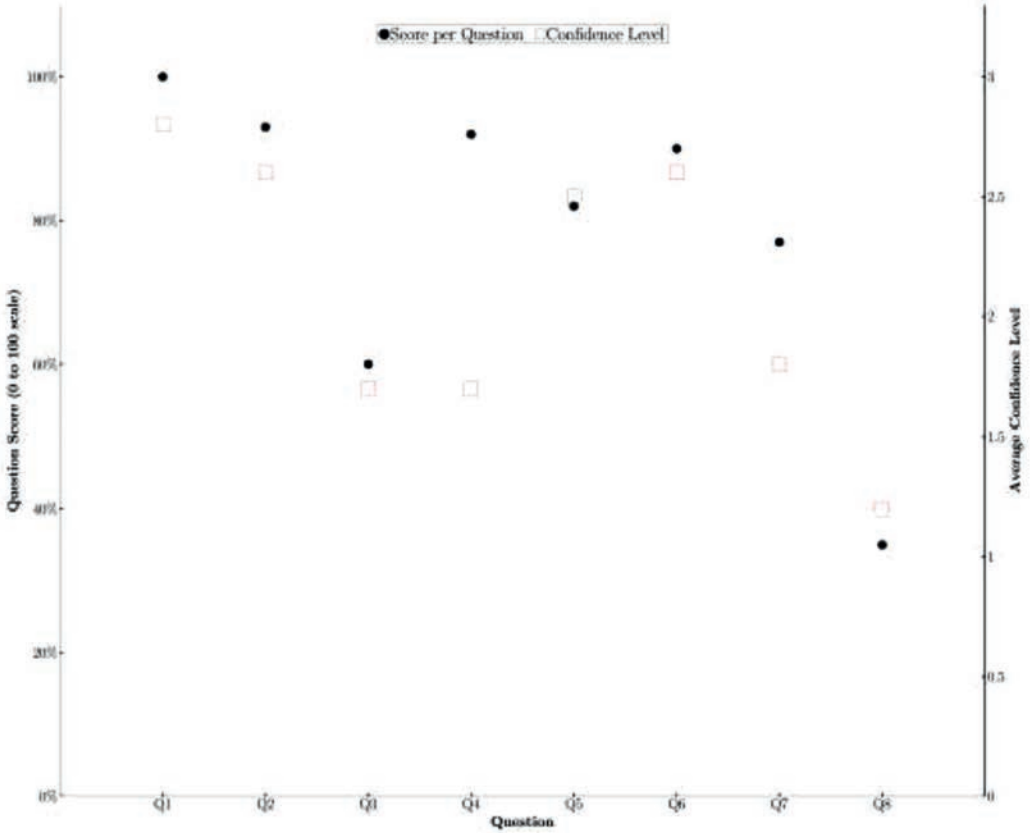


Figure 6. Course Assignment Average Grade.

The homework with the best performance in both groups was HW 3. In HW 3 students participating in "ungrading" had an average grade of 95 with the lowest grade of 88 whereas the students who did not participate had an average of 96 and the lowest grade earned was 86. In HW 4, the average slightly decreased for students participating in "ungrading" with an average of 96 and the lowest grade of 0. On the other hand, students who did not participate in "ungrading" had a low average of 55 and a minimum grade of 0. Both groups had students not turning in their HW, as a result, the grade average was affected. In general, it can be observed that students' participation in "ungrading" improved their grades as the semester evolved. Whereas the other group of students maintains a more variable performance.

Final Exam Performance

This section considered if utilizing "ungrading" practices on the midterm exam made an impact on the student's overall course performance. To do so, a question similar to question 8 on the midterm exam was also incorporated in the final exam. This strategy was made with the purpose of evaluating student's learning and material retention. Question 8 on the midterm exam read as follows:

"The back tangent for a simple horizontal curve has a bearing of N40°W. The PI of the curve is located at station 238+12.88 and has been assigned coordinates of: Northing: 10000; Easting 10000. The curve has a radius of 2,400 feet and Δ of 25° Left. The 2-lane road has 12 feet wide lanes, a design speed of 55 MPH, and an e max of 6%. What is the cross slope of the road at station 232+00?"

Similarly, question 8 on the final exam was stated as follows:

"The back tangent for a simple horizontal curve has a bearing of S25°E. The PC of the curve is located at station 44+72.12 and has been assigned coordinates of: Northing: 1000; Easting 1000 and has a centerline elevation of 393.55 feet. The curve has a radius of 3,000 feet and Δ of 35° Left. The 2-lane road has 11 feet wide lanes, a design speed of 65 MPH, and an emax of 6%. This section of road has a constant –1.1% longitudinal grade. What is the elevation of the inside and outside edge of pavement at station 63+80?"

The results of the average scores of the students participating in the "self-assessment" process versus the students opting not to participate are displayed in Figure 7. Both groups showed an improvement in their performance. Students participating in the "self-assessment" had initially earned an average of 35% of the total score they could earn in question 8, whereas on the final exam, these

students improved their score with an average of 96%. Students deciding not to participate in the "self-assessment" also presented an improvement in their grades. Initially, their average of the total score they earned on question 8 in midterm was 21%, whereas on the final exam, these students improved their score with an average of 79%. Even though the group experienced an increase in their score, the students who participated in "self-assessment" had a better performance.

With regards to the confidence level, both groups of students had a higher confidence in comparison to the actual score earned in their midterm exam. The opposite trend is observed on the final exam. The student's confidence level was lower in comparison to the grade they earned. These results are correlated to the trends observed in the confidence levels and actual scores for Q1, Q2, and Q6 (see Figure 5). Therefore, the student score was related to their confidence in solving the problem, in other words, when the confidence level is high, grades are high.

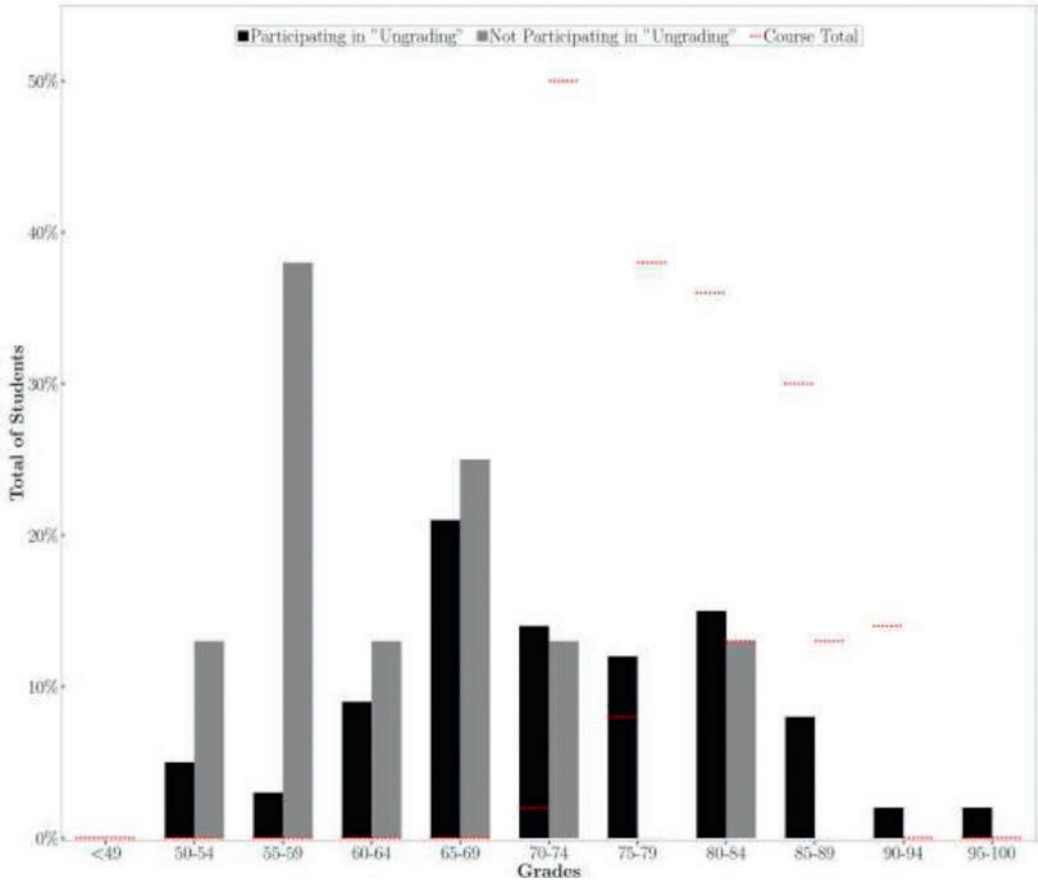


Figure 7. Students' Confident Level and Score on Question 8 of the Midterm and Final Exam.

Overall Course Performance

To further evaluate the effectiveness of utilizing "ungrading" practices we created a histogram (See Figure 8) with the results from the midterm exam and overall course grade. The results were tabulated by grouping the grades in increments of five and these were evaluated for both the group of students who participated in "ungrading" and for the ones who did not.

The grades were higher for the students who participated in the self-assessment process than those who did not. The results from the students who did not participate in "ungrading" show that their midterm grades were lower because the majority of them were below the 70-74 score range (though the sample was limited to eight students who choose not to participate in the self-assessment). On the other hand, students that participated in "ungrading" had a more normal distribution of their grades.

The results shown in Figure 8 can be used to support the effectiveness of the "ungrading" approach in that it did help students improve their course performance and therefore, obtain a better

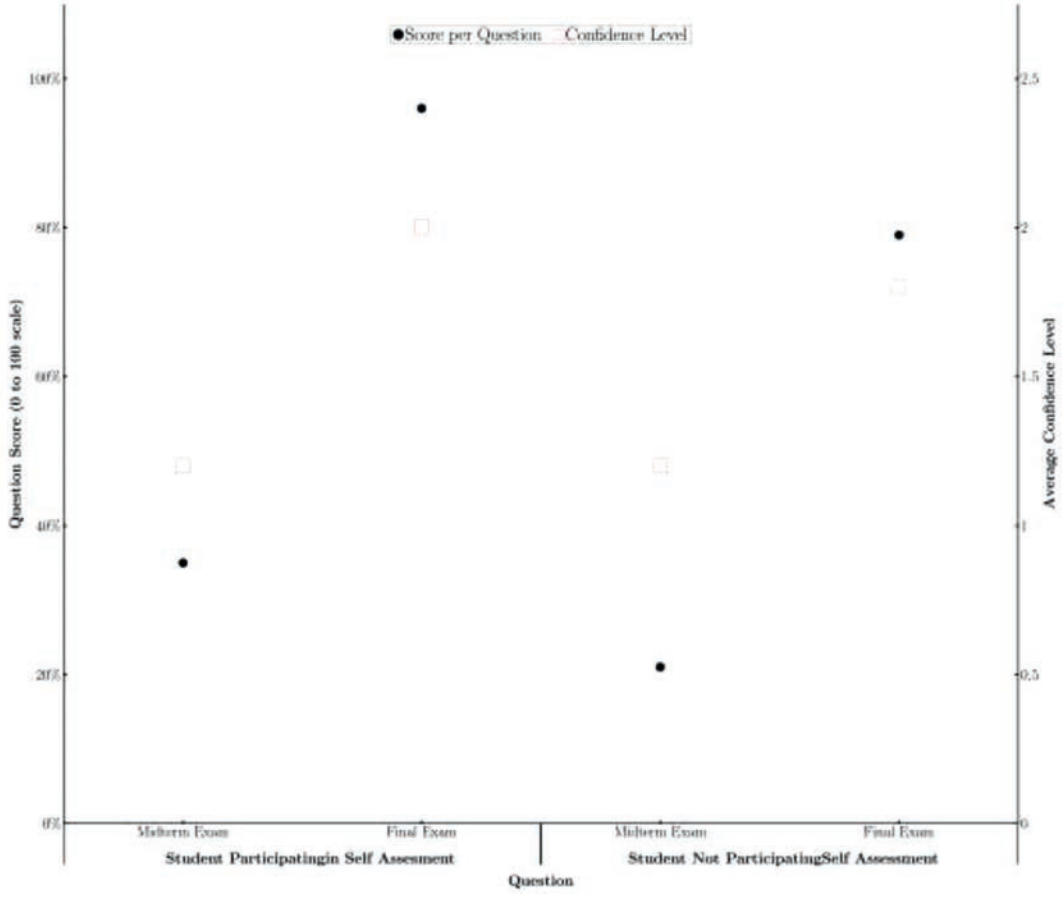


Figure 8. Course Performance.

overall course grade. However, the self-selection of the participating/non-participating students could also have played a role in the results because students who are more engaged in the classroom activities and processes are likely to score higher in the course.

Student's Feedback

At the end of each semester, students are asked to voluntarily provide feedback about the strengths and weaknesses of the course and how the course can be improved. The instructor specifically asked the students to provide feedback about the self-assessment in the free-text questions on the evaluation. The instructor received a copy of this feedback after the final grades were posted. This feedback was then reviewed to learn about students' experiences in the course and their experience with regard to their participation in "ungrading" for the midterm exam.

Our goal was to evaluate trends in students' feedback and identify how this impacted their learning. A total of six students provided comments about "ungrading" in the overall course evaluation process at the end of the semester. The students providing feedback considered "ungrading" to be good for them to learn the material. However, students also mentioned that at the time they were navigating through the process, there were a lot of uncertainties that instead of helping them, made them feel more stressed. These findings indicate that there is still room to improve and more specifications need to be added to help students on the adaptation of the ungrading practices.

CONCLUSION

Multiple studies examined the effectiveness of new pedagogical approaches that promote the reduction or elimination of a traditional grading system. However, the literature on "ungrading" practices in STEM courses is limited and the existing literature does not address the effectiveness of this method in STEM major courses in which student enrollment is greater than 60. This research contributes to this attempt by incorporating self-assessment into an undergraduate civil engineering course with a total enrollment of 72 students.

Ungrading was implemented only for the midterm examination and students' participation was optional (though 60 of the 68 students who consented to be part of the study also participated in the self-assessment process). To monitor student performance and the effectiveness of "ungrading," the overall course performance and student feedback were evaluated to see if there was a difference in grading between students who opted to participate and those who did not. This is a

similar finding found in students' feedback, where they indicated that they generally find "ungrading" beneficial to their learning.

Our results also indicated that it is crucial to repeatedly explain the idea behind "ungrading" so that students are aware of what is expected of them and why this is so crucial for their learning. Students participating in this study indicated that navigating the process was confusing, therefore, the rules and directions on how to navigate the process need to be revised so that in the future, students do not experience confusion or stress when navigating the process. Failing to do only adds more stress and anxiety to students, which may counteract the benefits and the purpose of the "ungrading" approach.

The incorporation of "ungrading" should first be incorporated on small assignments to help students become familiar with the concept and learn what is expected from them during the evaluation period. Students need a transition period that allows them to become familiar with the process, especially since there is a reluctance to the "ungrading" approach because students in STEM majors are strongly driven by a mentality that engages them in a course with a more test-focus approach rather than a learning approach (Gorichanaz 2022). There is still room to grow and many more strategies to implement. These initial findings on our effort to incorporate "ungrading" into an engineering program were not perfect but the results are promising.

Limitations

Building trust between students and the instructor is linked to students' perception of the effectiveness of learning in "ungrading." Therefore, it might affect the results of this study. At the time this study was performed, only the midterm examination utilized "ungrading," and since this is only one portion of their overall grade, students did not have enough time to learn how to properly work with this approach let alone build trust with the instructor. In the future, it would be ideal to do "ungrading" for course assignments so that the instructor builds trust with the students so that by the time the midterm is performed, students can engage better with the "ungrading" review process. Department or university-wide approaches and applications of "ungrading" may also prove useful for building students' trust in the process.

In addition, this study is limited to a single sample size and linked to a single institution. To have a better understanding of the effectiveness of "ungrading," a larger sample size is needed. It is recommended to consider graduate and undergraduate level courses in the STEM field. Another important factor to consider for future studies is the evaluation of learning performance based on different diversity, equity, and inclusion backgrounds. Additionally, it would be useful to evaluate if "ungrading" indeed promotes a more inclusive learning approach and quantify this impact.

REFERENCES

American Institute of Stress (2019) Stress: An epidemic among college students. (https://www.stress.org/stress-an-epidemic-among-college-students) Retrieved 1/17/2023.

Ashby-King D. (2021) More Than Just a Variable: COVID-19 and the Call to Complicate Communication Education Research. *Communication Education* 70(2): 205–7.

Aslanian S and Roth A (2021) Under pressure: Inside the college mental health crisis. APM Reports, 19 August. (https://www.apmreports.org/episode/2021/08/19/under-pressure-the-college-mental-health-crisis) Retrieved 1/17/2023.

ATL (2020). "Ungrading." Academy for Teaching and Learning (ATL), *Baylor University*. 2020. (https://www.baylor.edu/atl/index.php?id=984862) Retrieved 1/7/2023.

Bouchrika, I. (2020) 50 Current Students Stress Statistics: 2020/2021 *Data, Analysis, & Predictions.* (https://research.com/education/student-stress-statistics) Retrieved 1/17/2023.

Burke, L. (2020) "#PassFailNation." *Inside Higher Ed.* (https://www.insidehighered.com/news/2020/03/19/colleges-go-passfail-address-coronavirus) Retrieved 1/17/2023.

Bloodgood R., Short J., Jackson J., (2009) A Change to Pass/Fail Grading in the First Two Years at One Medical School Results in Improved Psychological Well-being. *Academic Medicine* 84(5): 655–62.

Blum, S. D. (2020). Ungrading: Why rating students undermines learning (and what to do instead). *West Virginia University Press*, 1st Edition.

Brennan, J., and Magness, P. (2019). Cracks in the Ivory Tower: The Moral Mess of Higher Education. *Oxford University Press*.

CITLS (2022). What Is Ungrading? Center for the Integration of Teaching, Learning, and Scholarship (CITLS), Lafayette College. (https://citls.lafayette.edu/what-is-ungrading/.

Dosmar, E. and Williams, J. (2022). Ungrading Assessment Practices. *The National Teaching & Learning Forum*, 31: 1-3. https://doi.org/10.1002/ntlf.30333.

Elbow, P. (1993) Ranking, Evaluating and Liking: Sorting Out Three Forms of Judgment. *College English* 55(2):187–206.

Espinola, M. (2018) History of the College Grading Scale 2018. (https://gradehub.com/blog/college-grading-scale/#:~:text=The%20first%20college%20grading%20scale,serving%20as%20a%20passing%20mark) Retrieved 1/7/2023.

Eyler, J. (2018) How Humans Learn: The Science and Stories Behind Effective College Teaching. Morgantown, WV: *West Virginia University Press*.

Flaherty, C. (2019) When Grading Less is More. *Inside Higher Education.* April 19. (https://www.insidehighered.com/news/2019/04/02/professors-reflections-their-experiences-ungrading-spark-renewed-interest-student) Retrieved 1/10/2023.

Gilley, B., and Clarkston, B. (2014). Collaborative Testing: Evidence of Learning in a Controlled in-class Study of Undergraduate Students. *Journal of College Science Teaching* 43(3), 83–91.

Ginexi, E. (2003) General Psychology Course Evaluations: Differential Survey Response by Expected Grade. *Teaching of Psychology* 30, no. 3: 248–51.

Goldrick-Rab, S. (2021) Students are Humans First: Advancing Basic Needs Security in the Wake of the COVID-19 Pandemic. *Journal of Postsecondary Student Success* 1(1): 3–17.

Gorichanaz, T. (2022) It made me feel like it was okay to be wrong: Student experiences with ungrading. Active Learning in Higher Education, 0(0). https://doi.org/10.1177/14697874221093640

Guberman, D. (2021). Student Perceptions of an Online Ungraded Course. *Teaching & Learning Inquiry* 9 no. 1. http://dx.doi.org/10.20343/teachlearninqu.9.1.8

Greenberg, K., Sohn, B., and Moret, L. (2022) Life in an Ungraded Course, *College Teaching*, DOI: 10.1080/87567555.2022.2046998.

Jarvis, C. (2020) Testing an 'Ungrading' Approach. *Chemical & Engineering News.* **2020**, 98(16), pp 20-23. (https://pubs.acs.org/doi/full/10.1021/cen-09816-feature1) Retrieved 1/7/2023.

Jennings, S. (2021). Ungrading: A Review, a Retrospective, a Messy Path Ahead, Digital Instruction Support Community, College of Education, *Michigan State University.* (https://education.msu.edu/digital-instruction-support-community/uncategorized/ungrading-a-review-a-retrospective-a-messy-path-ahead/) Retrieved 1/7/2023.

Koehler, A.A., Meech, S. (2022). Ungrading Learner Participation in a Student-Centered Learning Experience. *TechTrends* 66, 78–89. https://doi.org/10.1007/s11528-021-00682-w

Kohn (1993) Punished by rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes. Boston, MA: *Mariner Books*.

Kohn, A. (2018) Punished by rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes. Boston, MA: *Mariner Books*. (Original work published 1993).

Milton, O., Pollio, H., and Eison, J. (1986) Making Sense of College Grades: Why the Grading System Does Not Work and What Can be Done About It. San Francisco, CA: Jossey-Bass.

Open Education (2020). "Ungrading STEM" Open Education at University of Alaska Fairbanks. (https://open.uaf.edu/ungrading-stem/). Retrieved 10/19/2023.

Pascoe, M., Hetrick, S., and Parker, A. (2020) The Impact of Stress On Students in Secondary School and Higher Education. *International Journal of Adolescence and Youth* 25(1): 104–12.Pulfrey et al., 2011).

Stančić, M. (2021). Peer Assessment as a Learning and Self-assessment Tool: A Look inside the Black Box. *Assessment & Evaluation in Higher Education*, 46(6), 852–864 https://doi-org.ezproxy.lib.purdue.edu/10.1080/02602938.2020.1828267

Sharp, J. (1997) Ungrading: Adding Learning Intensive Writing Assignments Without Increasing Grading Load. Paper presented at *1997 Annual Conference*, Milwaukee, Wisconsin. 10.18260/1-2--6851.

Supiano, B. (2019) Grades Can Hinder Learning. What Should Professors Use Instead? *The Chronicle of Higher Education*, July 19. https://chronicle.com.

Supiano, B. (2022). The unintended consequences of 'ungrading': Does getting rid of grades make things worse for disadvantaged students? *The Chronicle of Higher Education*, (https://proxying.lib.ncsu.edu/index.php/login?url=https://www.proquest.com/trade-journals/unintended-consequences-ungrading/docview/2675461716/se-2) Retrieved 1/7/2023.

Schinske, J. and Tanner, K. (2014) Teaching More by Grading Less (or Differently). *CBE Life Sciences Education* 13 (2):159–66. doi:10.1187/cbe.CBE-14-03-0054.

Schneider, J. and Hutt, E. (2014) Making the Grade: A history of the A–F Marking Scheme. *Journal of Curriculum Studies* 46(2): 201–24.

Stommel, J. (2020). How to Ungrade. Jesse Stommel. Jesse Stommel. (https://www.jessestommel.com/how-to-ungrade/) Retrieved 1/17/2023.

Stommel (2021) Grades are Dehumanizing; Ungrading is No Simple Solution. Jesse Stommel. (https://www.jessestommel.com/grades-are-dehumanizing-ungrading-is-no-simple-solution/) Retrieved 1/10/2023.

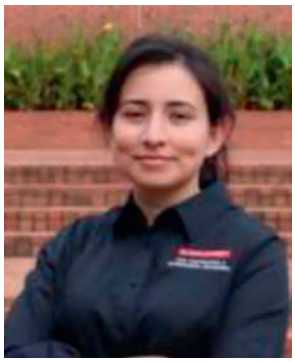
Spring L, Robillard D, Gehlbach L, (2011) Impact of pass/fail grading on medical students' well-being and academic outcomes. M*edical Education* 45(9): 867–77.

Talbert, R. (2022). "A Professor Shares the Benefits and Drawbacks of Ungrading (Opinion)." Inside Higher Ed., Higher Education News, Events and Jobs. (https://www.insidehighered.com/advice/2022/04/27/professor-shares-benefits-and-drawbacks-ungrading-opinion). Retrieved 10/19/2023.

Veletsianos, G. and Houlden, S. (2020) Radical Flexibility and Relationality as Responses to Education in Times of Crisis. *Postdigital Science and Education* 2: 849–62.

Wettergreen, S., Brunner, J., Linnebur, S., Borgelt, L., and Saseen, J. (2018). Comparison of Faculty Assessment and Students' Self-assessment of Performance During Clinical Case Discussions in a Pharmacotherapy Capstone Course. *Medical Teacher*, 40(2), 193–198.

AUTHORS



Dr. Minerva Bonilla is an assistant professor at Texas A&M University. At the time of this research, Dr. Bonilla was a Ph.D. candidate from the Civil, Construction, and Environmental Engineering program at North Carolina State University. Her research interest is in Construction and Transportation Infrastructure. She specializes in areas of constructability, modern unconventional intersections and interchanges, and funding for transportation infrastructure. Her current research efforts are related to the identification of construction issues affecting Alternative Intersections and Interchanges, equity in finding allocation and workforce development.

Dr. Daniel Findley, P.E., is an Associate Director with the Institute for Transportation Research and Education at NC State University in Raleigh, North Carolina. He specializes in economic impact analysis, multi-modal transportation, human behavior research, transportation engineering studies, asset management and inventory, and horizontal curve safety. He served as Principal Investigator (PI) or co-PI on over 80 funded projects and has conducted research for sponsors including NCDOT, FHWA, ACRP, ITE, and others. He holds a Ph.D. in Civil Engineering from North Carolina State University and is a licensed Professional Engineer (P.E.) in North Carolina. He also serves as an Adjunct Assistant Professor in the Department of Civil, Construction, and Environmental Engineering at NC State University.