2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #37531

WIP: Implementing an Alternative Grading Scheme in a Large Enrollment Differential Equations Course: Lessons Learned

Dr. Hadas Ritz, Cornell University

Hadas Ritz is a senior lecturer in Mechanical and Aerospace Engineering, and a Faculty Teaching Fellow at the James McCormick Family Teaching Excellence Institute (MTEI) at Cornell University, where she received her PhD in Mechanical Engineering in 2008. Among other teaching awards, she received the 2021 ASEE National Outstanding Teaching Award.

WIP: Implementing an Alternative Grading Scheme in a Large Enrollment Differential Equations Course: Lessons Learned

Abstract

In this work in progress paper, we describe our efforts to implement a standards-based, mastery-based grading scheme in a large enrollment differential equations course. Alternative grading schemes have been implemented in a variety of ways (mastery grading, specification grading, standards grading, etc.) in a variety of contexts (subject, class size, class level, etc.) for many years and they have many proven benefits. Some of the main concerns with traditional grading schemes driving our attempt to implement an alternative grading scheme in our course include: the time and effort needed to grade homework, which is often not an accurate demonstration of individual effort and understanding; the use of partial credit to reward poor work with middling grades, which allows students to accumulate a decent course grade without ever demonstrating correct problem solving techniques; agonizing over the right number of points off for various mistakes.

Even though alternative grading schemes have been shown to address many of those concerns, there are still many barriers to implementing them, particularly in courses with large enrollment. Energized by a workshop run by the Mathematical Association of America (MAA) during summer 2022, we have undertaken to implement a standards-based and mastery-based grading scheme for a 500+ student differential equations for engineers course at this R1 university during fall 2022. While we have received generous assistance from many groups, challenges for implementation remain. In this paper we describe the context of the course and our efforts to overhaul the grading scheme, with an emphasis on challenges and lessons learned.

Course Structure

In this section we describe the course structure as it existed prior to the study semester, then the overall shape of the revised course, followed by details of the grading scheme.

Overall logistics

This four credit-hour Differential Equations course is taught fall and spring semester each year with semester enrollments in the neighborhood of 350-550 students. The target audience is engineering students, typically sophomores, but approximately a quarter of the enrolled students are in other colleges outside of engineering. Topics include generating and solving first order ordinary differential equations (ODEs); generating and solving second order linear ODEs; Euler integration of first and higher order ODEs; boundary value problems (BVPs); Fourier series; and introduction to partial differential equations (PDEs) through the heat equation, wave equation, and Laplace equation.

During the semester of this study there were two lecture sections with a combined enrollment of 509. The course staff consisted of two instructors, a head teaching assistant (TA), and seven additional graduate student TAs. Teaching technology use included: Canvas as the learning management software (LMS), used for posting documents, assignments, and announcements; Gradescope where students submitted most written work, and where TAs graded all quizzes and exams; WeBWorK[1], an online platform for automatically grading a large variety of math problems, for weekly practice problems and a limited number of "checkpoints" (described more fully below). This paper's author was the main architect of the grading scheme overhaul. Unfortunately, at the beginning of the semester the second instructor fell ill and was replaced several weeks into the semester. The substitute instructor had not been planning to teach a large course that term, and as such focused mainly on lectures and writing exam problems, but was not available for much additional help.

Students in the course attend 50-minute lectures three times a week, and a 50-minute TA-led discussion section once a week. In a typical semester (before the change in grading scheme) students would submit weekly graded homework consisting of textbook problems, take a "homework quiz" during the first 10-15 minutes of discussion section, take two preliminary (midterm) exams, and take a comprehensive final exam.

Changes for Fall 2022

The main goals of the new grading scheme were to: stop collecting and grading written homework; require correct answers for credit; give ample opportunity for reflection and feedback on mistakes. To accomplish these goals, we devised the following outline for the logistics of the course.

- Textbook practice problems were posted each week, and full solutions posted a few days later. These were never collected, but were available for students to practice.
- Weekly WeBWorK assignments were assigned. Students were required to submit these each week for participation credit. Students had unlimited attempts and received automatic feedback on correctness immediately.
- TA sections started with a quiz on the previous week's topic. This quiz was timed and collected and graded according to the two bit grading scheme described below.
- Midterm exams were still given twice during the semester, according to the grading scheme described below.
- Online assessments, called "checkpoints," were added, mostly as Canvas quizzes with some auto-graded problems and some hand-graded problems.
- The final exam was still given, and graded according to the grading scheme described below.

The only new aspects of the shape of the course were turning the textbook problems into ungraded practice problems, and adding checkpoints throughout the semester. The meat of the change had to do with individual assessment grading and course grades determination.

Learning outcomes and grading

The cornerstone of a standards-based grading scheme is to identify the critical learning outcomes (LOs) for a course. Table 1 shows the LOs we identified for this differential equations course. We chose not to include certain topics that are usually covered in the course, such as solving exact equations or solving PDEs with more complicated boundary conditions, because of the limited number of weekly quiz opportunities. Not listed is the "Effort" LO, which is designed to encourage participation in the WeBWorK assignments and in discussion section. This was essentially a free checkpoint for all students during this implementation.

Some of the LOs are related to topics covered early in the semester and for which we had many assessment opportunities (e.g. Separable and 2nd order). For those LOs students had to earn two checks to get credit towards their final course grade. Other LOs were covered later in the semester (e.g. Wave) or were more difficult to write assessment opportunities for (e.g. Modeling). For those LOs students only needed one check to earn credit. Checks could be earned on the weekly quizzes, on exam problems, or on checkpoints.

Each graded problem in the course was explicitly tied to a particular LO. Either the weekly quiz was on an announced topic, the checkpoint was explicitly associated with a LO, or exam problems were explicitly labeled with their relevant LO. A mastery grading scheme was used to grade each individual problem. Table 2 shows the possible score a student could earn on each problem. An important aspect of the course is that if students make a non-conceptual error on a problem, they can look over their work, identify the mistake, and submit a correction. These revisions were possible on all assessments except the final exam.

Final course grades were based on the number of LO checks earned, as detailed in Table 3. This table reflects changes made during the semester as it became clear that a sufficient number of checkpoints would not be available on time. Importantly, earning more than the required number of checks on a single LO did *not* affect students' grades. That means that if a student had already earned the required number of checks on a certain LO, or if they decided not to attempt a certain LO, they did not have to solve every exam problem. The midterm exams had 8 problems each, and the final exam had 19, but most students did not attempt more than 5-7 problems on a single exam (and they had the choice of problems to attempt).

Discussion

Successes

Some aspects of the course were very successful. The process of revision and resubmission of problems on which students earned a 2 led to some of the best office hours interactions we've ever had. This grading scheme strongly incentivizes students to understand and learn from their mistakes. Students can engage with the material on a deeper level instead of arguing over points.

Table 1: Learning outcomes for the course. Students had to demonstrate mastery of each LO a certain number of times to earn credit for that LO towards their final course grade.

LO	Description
Modeling	Generate differential equations to model physical problems and use the models to answer relevant questions.
Slope fields	Use slope fields to sketch differential equation solutions for given initial conditions (ICs).
Separable	Solve Initial Value Problems (IVPs) with separable ODEs by integration.
Integrating factor	Solve linear 1st order ODEs using integrating factor.
Equilibria	Identify stable and unstable equilibria for ODEs of the form $y' = f(y)$ and plot qualitative sketch of solutions.
Numeric	Use Euler method for numerical integration of 1st order ODEs.
2nd order	Solve homogeneous linear constant coefficient 2nd order ODEs using the characteristic equation for real, complex, and repeated roots.
Undetermined coefficients	Use method of undetermined coefficients to find particular solutions to nonhomogeneous 2nd order constant coefficient ODEs.
Reduction of order	Use reduction of order to find linearly independent solutions to linear ODEs.
Technique ID	Identify which solution techniques can be used to solve given ODEs.
Higher order numeric	Euler integration for higher order ODEs.
Vibrations	Write equations to model systems with free and forced vibration of undamped and damped mass-spring-damper systems; Solve such systems.
BVPs	Find eigenvalues and eigenfunctions of BVPs.
Fourier series	Find the coefficients in a Fourier series (FS).
FS Even/Odd	Identify whether a function will require a sine series, cosine series, or both.
Homogeneous heat	Use separation of variables to solve the homogeneous heat equation.
Non-homogeneous heat	Sketch solutions to the heat equation based on non-homogeneous boundary conditions (BCs).
Wave	Solve the wave equation based on given initial conditions.
Laplace	Solve steady state 2D heat conduction with simple BCs (Laplace equation).

Table 2: Possible scores on each problem in this mastery grading scheme. If a successful revision is submitted, a 2 can be upgraded to a 3.

Score	Description
3: Check received	This indicates a complete and correct solution up to and including the final answer. This is one "check" on the related LO.
2: Almost	Non-conceptual error such as an algebra error or misread number in the given problem, etc. Students can revise their solution and resubmit it.
1: Not yet	This indicates a student tried to solve the problem, but did not demonstrate understanding of the problem or solution methodology. A conceptual error exists. They have to try again on a new problem at the next opportunity.
0: Not attempted	This is just a flag to say that a student didn't try to answer a problem. If students are missing quizzes we want an easy way to notice so we can reach out and see what's going on.

Table 3: Mapping from number of LOs mastered to final course grade. For LOs for which two checks were required for mastery, earning one check was worth half a LO mastered.

# of LOs mastered	Course grade
12	D
12.5	$C-$
13	C
13.5	$C+$
14	$B-$
15	B
16	$B+$
17	$A-$
18	A
19	$A+$

Table 4: Questions pertaining to the grading scheme appearing on the end of semester course evals, along with student responses.

Questions	Student responses 1	2	3	4	5
This semester we piloted an alternative grading scheme based on mastery of learning outcomes [LOs] with revision attempts. We recognize that there were many logistical problems with the implementation. Please indicate your feelings about the alternative grading scheme. 1 = I disliked the idea from the start. 2 = I neither liked nor disliked the idea at the start but the execution was a detriment to my learning. 3 = I liked the idea at the start but the execution was a detriment to my learning. 4 = I neither liked nor disliked the idea at the start but ended up happy with the grading scheme. 5 = I liked the idea from the start and ended up happy with the grading scheme.	21	54	154	21	35
Assuming the logistics of the implementation can be fixed (timely checkpoints, timely grading, timely communication with students about progress, etc.), would you recommend we implement this grading scheme in future semesters? 1 = No, regular grading method is much better 5 = Yes, if the logistics are done right, the new scheme is much better for learning	61	22	39	55	106

We held extended weekend office hours right before exam revisions were due each time, and each time the room was filled with students digging into their mistakes to earn the checks.

On the end of semester course evaluations, we added two questions to the university form to ask about students' impressions of the revised grading scheme. The full questions and answers are shown in Table 4, but the most important aspect is that when asked if they recommend we implement the grading scheme in the future, over a third of respondents said the new scheme is much better for learning.

Anecdotal evidence also indicates students' support or at least open-mindedness towards this grading scheme, even though it is not something they have seen in other STEM courses at our university. We heard from several students that they learned much more in this math course than in any other, that they never looked at graded exams in other courses but they did in this one, that they appreciated the revision cycle. Even students who were very upset about the logistical failures, to the point that they were contacting the dean's office, admitted that they learned a lot in the course.

Challenges

While we were very excited about the idea behind the grading scheme, and while student interactions around the revision process were wonderful, there were many serious logistical challenges with our implementation. Challenges fall roughly into three categories: grading, checkpoints, automation.

Grading

A successful implementation of this mastery-based 4 point grading scheme requires: 1) grading to be consistent from student to student, from grader to grader, and from assignment to assignment; and 2) for revisions to be held to a consistent standard. We had trouble on both of those fronts. With 500+ quizzes to grade each week, plus up to 300+ revisions of the previous week's quiz, plus occasional exams or checkpoints to grade, including revisions, the grading load is enormous. To compensate for the large number of problems to be graded each week, the grading of each problem should be as efficient and uniform as possible. However, we did not provide sufficient oversight to the TAs and did not make a big enough point to the TAs or the students early in the semester just how good a solution has to be to earn a 2 instead of a 3. We wanted to give as many students as possible the opportunity to revise their work, but that diminished the "mastery" aspect of how we implemented the grading. There was also a lack of consistency between TAs as to what error earned a 1 versus a 2. In the future, either a single TA should grade all students' solutions to a single quiz or exam problem, or much more clear guidelines have to be given to TAs and enforced early in the semester.

We also did not sufficiently emphasize, to students or TAs, the importance of complete reflection in submitted revisions. We had a Canvas page that described in detail what a revision should consist of, including a fabricated example of work that received a 2 and how to properly revise for a 3. That Canvas page concluded with this sentence: "What you submit for the revision should include the original work with the highlighted error, the explanation, and the correct work." However, it wasn't until this author was grading revisions on a prelim problem that we realized that students were not in general following this format, and TAs were not in general enforcing this format. It is exactly this format that gets students to properly reflect on (and therefore learn from) their mistakes. Next time, it's important to make sure the TAs are enforcing this required format on submitted revisions, and rejecting revisions that fail to meet all aspects of the requirements.

Checkpoints

If the grading is going to be so exacting, and revisions will be rejected for failure to meet the reflection requirements, and only scores of 3 count towards course grades, then it's absolutely essential that a sufficient number of assessment attempts be available for each LO. Because of the large student-to-staff ratio, the only way to reasonably have a large number of assessments available is if some of them are automatically, instead of manually, graded. Checkpoints are designed to serve that purpose. However, it proved to be more difficult that expected to generate appropriate auto-graded problems for many LOs. Software does exist to generate questions banks with some randomization[2], however it primarily generates individual problems with solutions that must be graded by hand (though it does provide the solutions for each problem), which is not suitable for our needs. Using a combination of python and Matlab scripts generated by colleagues for their own unrelated course purposes, we were able to generate large questions banks for many LOs. These codes are limited to generating multiple choice questions at the current time. And for each LO, we had to come up with an algorithm with variable parameters to generate these banks of multiple choice questions. Some LOs (such as Equilibria) were well-suited to this version of question, but some (such as Undetermined Coefficients) were very poorly suited.

Checkpoints were finalized and released very late in the semester. They were delayed by time spent figuring out and repurposing the question bank generating scripts, and by time spent generating appropriate algorithms to generate randomizable questions. After release to the students, most checkpoints included one or more multiple choice questions plus a final question where students submitted their handwritten work supporting their multiple choice answers. The multiple choice questions were automatically graded, but a person had to look at each of the submitted handwritten solutions to make sure they supported the MC answers selected. We used Canvas SpeedGrader for this process, which we found cumbersome. Again, with 300+ submissions of most checkpoints, that was an extremely time-consuming process, and the timing ended up overlapping with exam writing and grading.

Automation

We grossly overestimated the extend to which tasks and processes associated with the course would be automated. For example, Canvas has a Learning Mastery gradebook, but because all of the quiz and exam grading happened in Gradescope, we did not manage to link those grades into the Learning Mastery gradebook in Canvas. We believe this should be possible with a python script through the Canvas API, but during the crush of the semester we were unable to get it figured out. The result being that we did not provide students with a timely update on their progress through the course, and in particular did not give them a sense of whether they were "behind" where they should be at certain points in the semester.

Another serious shortcoming in the automation was when students needed to submit revisions for checkpoints. Because the checkpoints were given primarily as Canvas quizzes, we had them submit revisions as comments to the original submission. This should nominally work, as Canvas can generate a daily notification email including a link to every comment submitted. Simply opening each link in that email in a separate tab allows a grader to look at each submitted revision. Unfortunately, if there are more than a certain number submitted in a single day, this notification system fails. That led to us missing dozens of revisions until well after the fact, and ever-changing instructions to the students about how to alert us. The sheer number of emails generated by this course was mind boggling, up to several dozen a day, each asking for individual attention and help submitting files.

Conclusion

Alternative grading schemes lead to many desirable outcomes, including increased student engagement with the material, training students to solve problems completely and correctly, and giving students the opportunity to recover from early poor performance. To be successful, such schemes must be implemented in a holistic way including many assessment opportunities throughout the course. The enormous amount of grading needed for a successful implementation is difficult to scale to large enrollment courses. In our first attempt, we identified several critical shortcomings and the solutions to those shortcomings going forward. Checkpoint opportunities must be created and deployed earlier in the semester. In the future we hope to write scripts to generate large question banks including different types of Canvas quiz capabilities (such as including images, multiple answer questions, numeric response). Mastery grading must be enforced more stringently and uniformly. We will make a point of training TAs more carefully early in the semester, and creating specific and uniform rubrics for quiz and exam grading. Revisions must be held to high standards with a strong emphasis on reflection so that students have the maximum opportunity to learn from their mistakes. Early in the semester we will be sure to reject revisions that don't meet the necessary standards. Full use must be made of the Canvas Learning Mastery gradebook, so that students have up-to-date information about their progress towards course objectives. We hope to develop scripts to connect Gradescope grades to Learning Mastery scores in Canvas through the API, though at this time we are unaware of other people making this automated connection.

Even though the logistics of implementing this mastery graded and standards based grading in our 500+ student course were very difficult, and our implementation fell far short of our goals, we remain enthusiastic about the possibility of improving the process and trying again. Students are open to this sort of grading scheme and the increased emphasis on learning over more traditional grading schemes. Once the implementation logistics have reached a smoother steady state we hope to assess the effects on students learning, possibly using the spring semester version of the course, taught by different instructors and without this grading scheme, as a control group.

Acknowledgments

This course revision attempt was made with the help and support of many colleagues. Thanks to the MAA Open Math Professional Development Workshop: "Redesigning Your Course for Mastery Grading" organizers Sharona Krinsky, Owynn Lancaster, and Robert Bosley. They ran an inspiring workshop. Thanks to head TA, Mark Walth, who went truly above and beyond in his duties. Brian Kirby and Dmitry Savransky provided their python and Matlab codes and their debugging help. Finally, Tara holm as chair of the Math department provided crucial additional TA resources towards the end of the semester when the full extent of the course trouble became clear, absorbed communications from many angry students (and parents), and never gave me a hard time about the trouble I caused. I'm sincerely grateful for her support and forbearance.

References

[1] The Mathematical Association of America, "What is webwork?" Available at https://webwork.maa.org/intro.html (2023/02/28).

[2] S. Clontz, "Checkit," Available at https://checkit.clontz.org (2023/02/28).