Mastery Grading in a Software Engineering Course

Carlos Rojas, San Jose State University

Carlos Rojas is an Assistant Professor in the Department of Computer Engineering at San Jose State University. His current research focus is on applying computational, machine learning, and visualization techniques to the understanding of biological mechanisms and systems. Dr. Rojas received his Ph.D., M.S., and B.S. degrees from UC Davis where his research focused on computational methods to compare 3D geometric shapes.

Prof. Gina M. Quan, San Jose State University

Gina Quan is an assistant professor in the Department of Physics and Astronomy at San Jose State University. She previously earned her Ph.D. in Physics Education Research at the University of Maryland, College Park. She studies student identity, community, and institutional change.

Mastery Grading in a Software Engineering Course

Abstract

Developing alternative forms of grading and assessment has been an area of ongoing interest in engineering education. Mastery grading is one form of alternative grading which allows students to have multiple attempts to demonstrate their progress toward learning objectives. While mastery grading has been implemented in a variety of contexts, it is less common in undergraduate engineering courses. This paper summarizes a faculty's first time implementation of mastery grading in an upper division software engineering course and presents a longitudinal study of students' perspectives toward mastery grading. We review the perspective of the faculty and students. We surveyed students throughout the semester to understand (1) the change in students' attitude towards mastery grading, (2) if students understood the mechanics of mastery grading and (3) the extent to which students found this system helpful for their learning versus traditional grading. We found that while students initially had a strong negative reaction towards the new grading scheme, over time, they seemed to prefer this grading method. Mastery grading made it possible to adapt to changes as issues arose. We then describe the mastery grading scheme implementation from the perspective of the instructor. Finally, we discuss the implications of this work for research on grading as well as the implications for instructors.

Introduction and Motivation

Traditional grading with a point system is time consuming, and it does not incentivize students to learn the material. As pointed out by Nilson [3], traditional grading can lead students to optimize for the numbers of points, and not learn the material. In this work, concerns about traditional grading motivated the instructor to reduce the time spent in grading and to incentivize students to focus on learning the content in an upper division software engineering course.

Mastery grading (also known in the literature as specifications grading) has been applied to a wide variety of fields [1,2,4,5]. In mastery grading, the course is built around assessments that are graded by a threshold that establishes mastery. There are rubric items that outline what it means to establish mastery for that assessment. There are no partial points, only a designation that lets the student know how well they mastered the assessment. Mastery of an assessment can be assigned a binary grade (pass/fail), but there are implementations where instructors have created up to four discrete levels of mastery. The assessments are then bundled and organized to create a mapping from the outcomes of the assessments to letter grades.

Mastery grading makes it very clear what grade a student has and what they need to do to improve their grade. The system also makes it easier for the instructor to grade because they only need to just check if the assignment met the criteria for mastery or not. By adding proper feedback, we can save time and improve learning, as shown by Pascal et al. [6]. However, there

were some challenges that the reader should know before converting their course (see Section Challenges).

Previous work in software engineering courses was limited. However, there were examples such as Ranalli and Moore [7] demonstrating what it is like to teach a mastery grading course for the first time. For learning about how mastery grading impacts student learning for a variety of engineering courses, the work by Schlemr and Vanasupa [5] is a prime example. Mastery grading can also be adapted to benefit hybrid courses [2]. In our work, we were inspired by previous study designs, but the implementation details were inspired by examples in Nilson [3].

Course Context

In Spring 2022, Rojas was the instructor of record for two sections of Software Engineering I using mastery grading (specifications grading). The sections had 83 students (34 and 49, respectively). The class was mostly Computer Engineers (CE) and Software Engineers (SE) with non-majors making up 10% and 30% in each section. The course took place Tuesday and Thursday for seventy-five minutes each session in the semester. This is one of the first upper division courses that students in CE and SE take. Both sections were given similar assignments with only minor details changed for exam questions.

Description of this Implementation of Mastery Grading

The most important part of mastery grading is deciding how to organize course assessments. In this course, Rojas created five assessment categories: quizzes, homeworks, exams, a project, and individual project contributions. Given that Rojas taught this course several times before, he assumed that previous assignments already mapped well to the Course Learning Objectives (CLOs) and he did not need to update the assignment content. Next, Rojas converted each assignment so that he could grade them as either needs revision (NR), low pass (LP), or high pass (HP) with the only exception being the exam category, which was graded traditionally. Each assignment's rubric varied slightly, but for a student to earn a LP they would need to answer all the questions deemed necessary to show mastery in the assessment. To earn a HP the student would need to meet the requirements for LP and correctly answer one or more advanced questions.

There was no partial credit for any of the assignments. Students that did not meet the LP or HP thresholds on assignments could revise and resubmit them. This was the main mechanism that was intended to incentivize students to revise subjects they had not mastered yet. Rojas imposed limits on the number of retakes to reduce students' ability to brute force the solutions. Of course, this also sets boundaries on the amount of grading.

Rojas created a currency, called tokens, that students could use to add flexibility to their learning. For example, tokens could revise an assignment, or extend a deadline by 24 hours. Students earned tokens via short assignments, which were not related to the CLOs, but information that would be helpful to students. For example, one assignment was to learn about growth mindset and another was on the Pomodoro technique. Future work is needed to understand the impact of these assignments on student outcomes.

Revisions and resubmits had two variations. For homeworks, Rojas allowed students to fix their mistakes and resubmit the same assignments. For quizzes, Rojas had students attempt a new quiz with new questions, but on the same topics as the previous quiz.

Given the tertiary grading system, Rojas developed a mapping to traditional grades, which is shown in Table 1. However, the number of assignments was reduced from 5 to 3 as the semester went on because of the number of revisions. Note that Exam Grade is a percentage; Rojas graded the exams (midterm and final) traditionally and there were no revisions for them. However, the exam category was the maximum between the midterm and final. For a student to earn a letter grade, they would need to meet all the specifications for that letter grade. Meaning that the lowest grade in a category would dictate their grade.

Grade	***Quizzes***	***Homework***	***Exam Grade***	***Project***	***Project Contribution***
A plus	5 High Pass	5 High Pass	*>=95%*	4 High Pass	4 High Pass
A	5 High Pass	5 High Pass	*>=90%*	4 High Pass	4 High Pass
A minus	4 High Pass, 1 Low Pass	4 High Pass, 1 Low Pass	*>=90%*	4 High Pass	4 High Pass
B plus	3 High Pass, 2 Low Pass	3 High Pass, 2 Low Pass	*>=85%*	3 High Pass, 1 Low Pass	3 High Pass, 1 Low Pass
B	3 High Pass, 2 Low Pass	3 High Pass, 2 Low Pass	*>=80%*	3 High Pass, 1 Low Pass	3 High Pass, 1 Low Pass
B minus	2 High Pass, 3 Low Pass	2 High Pass, 3 Low Pass	*>=80%*	2 High Pass, 2 Low Pass	2 High Pass, 2 Low Pass
C plus	1 High Pass, 4 Low Pass	1 High Pass, 4 Low Pass	*>=75%*	1 High Pass, 3 Low Pass	1 High Pass, 3 Low Pass
C	1 High Pass, 4 Low Pass	1 High Pass, 4 Low Pass	*>=70%*	1 High Pass, 3 Low Pass	1 High Pass, 3 Low Pass
C minus	5 Low Pass	5 Low Pass	*>=70%*	4 Low Pass	4 Low Pass
D plus	4 Low Pass	4 Low Pass	*>=65%*	3 Low Pass	3 Low Pass
D	4 Low Pass	4 Low Pass	*>=60%*	3 Low Pass	3 Low Pass
D minus	3 Low Pass	3 Low Pass	*>=60%*	2 Low Pass	2 Low Pass
F	Specifications for D minus not met				

Table 1: Letter grade conversion for mastery grading. There would be five quizzes and five

homework assignments. A total of four project milestones. The lowest grade in a category would be the letter grade.

Methods

Our data collection served the purposes of (1) ongoing assessment of implementing mastery grading, in order to inform future iterations to the system and (2) documenting the perspectives of students and the instructor to understand how attitudes toward mastery grading evolve over time. Toward these ends, we collected surveys, documents and artifacts, and instructor reflections.

As Quan was not the instructor of record for the course, she led the collection of survey data. Five identical surveys were sent to all students every three weeks in both sections of the course. The surveys were a combination of Likert scale and free response questions. Questions broadly covered students' understanding of mastery grading (e.g., "I understand how I can improve my grade on a project or homework assignment" [Strongly Agree to Strongly Disagree]), students' satisfaction with mastery grading (e.g., "I prefer mastery grading to other forms of assessment." [Strongly Agree to Strongly Disagree]), and the extent to which mastery grading was reflective of their learning (e.g., "To what extent do your grades on the project and homework assignments reflect how much you know? Explain." [Short Answer]). While some students completed the surveys anonymously, students who completed at least three of the five surveys and indicated their name were compensated with a $10 Amazon eGift card. Of the 83 students in the course, the number of completed surveys ranged from 12 to 18 participants, and we had 69 complete surveys throughout the semester. Rojas did not have access to the survey data until the end of the semester, but Quan occasionally shared broad patterns as formative feedback during the semester.

To capture the instructor's perspectives on the course as well as how the implementation of mastery grading shifted over time, Rojas engaged in regular reflective journaling. We also collected documents and artifacts associated with the course including emails to and from students which discussed mastery grading and syllabi from the focal semester and previous semesters. We also viewed student course evaluations administered by the university (N=59) to help us contextualize our interpretations of the survey data.

Our analytical process began by looking for patterns over time in the Likert scale items. Because of the limited sample size in each survey, we were not able to draw broad generalizations from the data. However, we were able to identify some patterns in students' perspectives over time that could be further explored in a future study. As we noticed patterns in the Likert scale data, we used the short-answer responses to help us understand some of the classroom experiences leading to particular ratings. As Rojas was the course instructor, he also shared insight on course modifications that may have led to certain changes over time. Though student course evaluations

were not designed with our specific goals in mind, the larger participation rate added additional perspective as we weighed particular interpretations of our survey data.

Results from Instructor Perspective

It took several assignments for the students to fully understand the system. The positive outcomes were that there were fewer questions about requests for grade changes due partial credit. The requests weren't completely eliminated because there were still students that would earn a low pass due to missing one question and then would ask if that could be pushed to a high pass. This again was eventually resolved by improving the thresholds for high pass and low pass.

Grading was easier with the three tier grading system because partial credit wasn't a factor. The teaching assistant mentioned that they had a harder time grading the midterm and final because judging partial points was time consuming.

There was an increase in the average grade point from 2.8 to 3.1 from the previous two semesters (fall 21, spring 21) that were taught with traditional grades. In the mastery grading sections, the number of assignments were similar to previous semesters, but due to revisions the workload was higher.

Analysis and Results from Student Perspective

Patterns Likert-Scale Data

We began analyzing student survey responses by first generating plots of the distribution of responses to each Likert item (Figure 1). This helped us infer the relative level of agreement across items. For example, Figure 1 represents the Likert items from the first survey administered in the class. This data suggests that most respondents felt they had the background knowledge and resources needed to succeed in the class, and had an understanding of how to improve their grade. However, the median respondent did not prefer mastery grading over other forms of assessment.

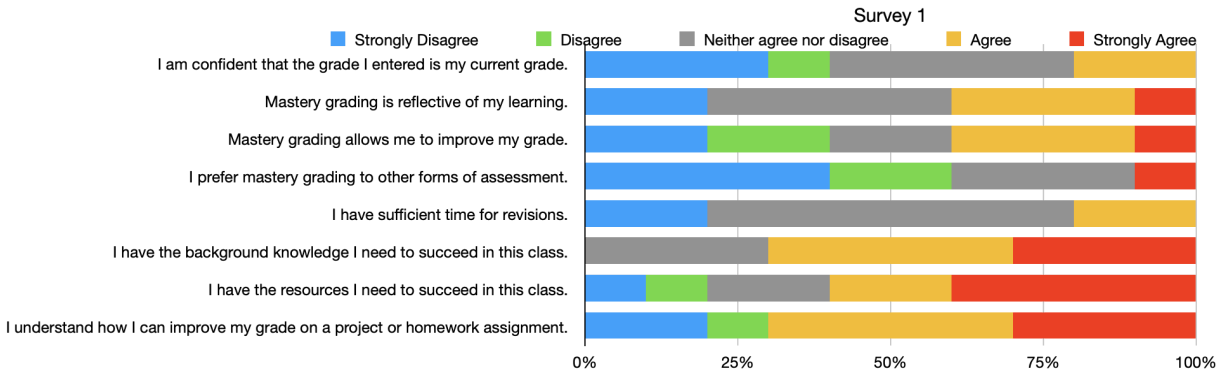

Figure 1: Distribution of student responses on Survey 1 (N=10)

We generated these plots for every survey. The results of Survey 5, which was given at the end of the semester, are shown in Figure 2. Comparing Figures 1 and 2 we see that responses to Survey 5, on the whole, were more likely to "Agree" on most items.

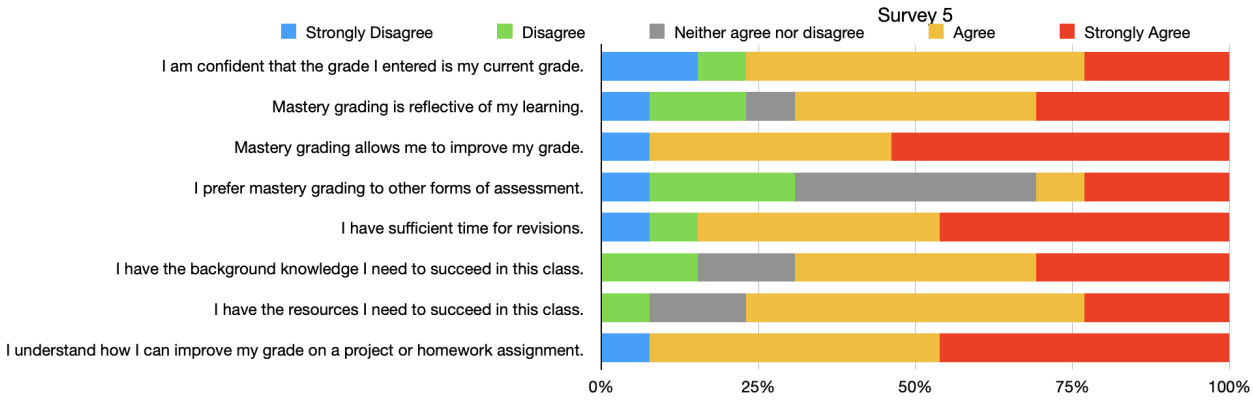

Figure 2: Distribution of student responses on Survey 5 (N=13)

Next, to support us in comparing broad patterns over time, we converted the Likert-Scale items to an interval scale (2=Strongly Agree, 1=Agree, 0=Neither Agree nor Disagree, -1=Disagree, -2=Strongly Disagree) which is shown in Figure 3. Treating the data as interval does have limitations and loses information about the spread and distribution of responses, but it does allow us to easily see change over time in the class.

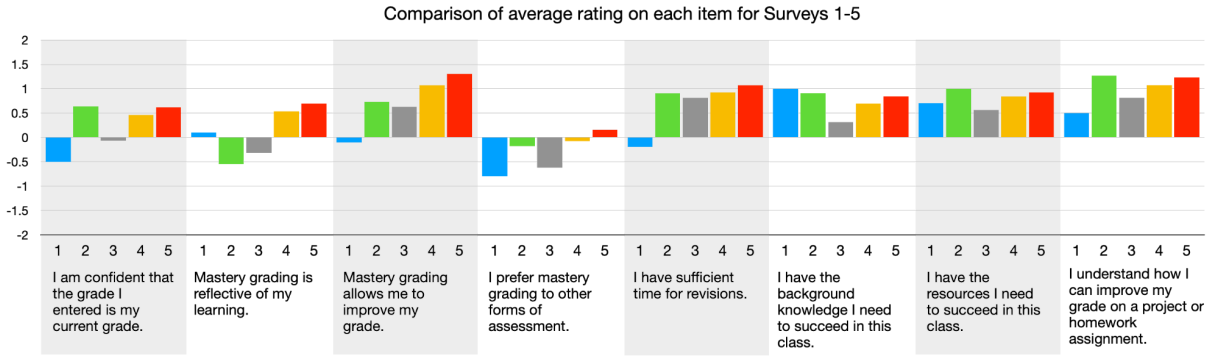

Figure 3: Results of survey questions taken at five time points. We converted the likert scale to range from -2 to +2.

We then looked at the free-response items to help us gain more context for the patterns in Likert-Scale items. We now present three themes that emerged.

Initial confusion about course grading scheme, with frustration decreasing over time

Consistent with challenges identified by the instructor, students' initial confusion and frustration with the course grading scheme was reflected in the surveys. For example, participants were asked to estimate their current grade in the class and then rate their confidence in the grade they entered (Item 1). Participants did tend to be more confident over time, which is perhaps unsurprising given that students were given more feedback over time. Additionally, participants

also seemed to be more likely to see mastery grading as reflective of their learning over time, with a dip around the middle of the semester (Item 2). Within the free response, we did see evidence of this confusion initially. As an example of one student response expressing confusion: “As of this survey, I cannot state how much projects and homework assignments have reflected the grade, considering I do not have a grade in the class yet”—*Survey 1* and frustration: “So far the rubrics are incredibly punishing and demoralizing, but I dont know any results.”—*Survey 1*.

By the end of the semester, students did feel more confident in what their grades would be, and in general felt that their grades were reflective of their learning. One student who initially answered “Disagree” to Items 1 and 2 later wrote “Honestly, it's not too bad now. We have lots of time to revise assignments and quizzes which makes me feel more confident in my final grade.”—*Survey 5*. Similarly another student wrote “My grades accurately reflect my knowledge. The more work I put into the class the better my grades get. ”—*Survey 5.* Overall, survey responses suggest that while the course grading scheme was initially confusing, and the cause of student frustration, responses over time improved.

Dissatisfaction with Weighting of Mistakes

Another theme we identified was tied to the initial dissatisfaction with grading (described above), but was specifically tied to the impact of mistakes. As the mastery grading process was firm on assigning some scores as “Need Revision” some students expressed frustration with being given assigned grades. These responses were most prominent in Surveys 2 and 3 (perhaps consistent with the dip in Item 2 for these surveys). For example, one student wrote: “At this point, I don't believe they reflect how much I know due to how mastery grading works on catering to perfection. While I appreciate the opportunity to get a better score, losing all the points for a single mistake on homework feels diminishing.”—*Survey 2.* And in a similar vein, “I don’t like that we don’t get partial credit, because I know more than just deserving a zero, but I might not always know everything to get 100%”—*Survey 3.* From these responses, we see that some students were more likely to see “Needs Revision” as a demoralizing “0” rather than an opportunity for improvement.

As we see in both of those quotes above, students often saw the “Needs Revision” to be a result of what they felt was a minor gap in knowledge, and this was (in their view) not indicative of their broader capabilities. Another student wrote: “Not a lot, because sometimes a silly mistake costs me a high pass.”—*Survey 3*. Based on the survey data alone, it is unclear whether these grades were the result of what was truly a “silly mistake” compared to a deeper conceptual misunderstanding. We see both possibilities as plausible—an individual grader may have misinterpreted a minor syntax error as a deeper misunderstanding. On the other hand, a student may have had a fundamental misunderstanding of a topic and may have only viewed the problem as a “silly mistake.” In either case, this suggests a need for more thorough dialogue with students about the relative importance of different types of errors.

Revisions, Learning and Improvement

In contrast to the previous theme, other students spoke to the importance of revisions in supporting their learning. One student wrote: "I feel that because of the revisions were allowed to take, it is a reflection of how much I've learned in the class rather than how much I remembered on the given day of a quiz"—*Survey 3*. This response suggests that the student felt the grade overall ended up more accurate. Similarly one student wrote: "Because of the way our professor grades things, a couple of really small mistakes can end up making you fail the assignment and in turn have a failing grade for the class, but revisions allow us to correct those so it's pretty accurate in the end I think."—*Survey 3*. In this response, we see a similar frustration with the impact of mistakes described above, but this student also found the revisions to allow them to improve their grade. This is consistent with the survey responses to Item 3, which suggests that over time students were in more agreement that mastery grading allowed them to improve their grade.

Other student responses also suggested that the ability to do revisions also impacted how students engaged with the class. One student wrote: "They help show which sections we are struggling with and allow us to see what we should depend on more time practicing."—*Survey 2*. Similarly, another wrote "It has encouraged me to look more closely at my past work and make adjustments to my understanding before continuing on to other topics."—*Survey 4*. These responses suggest one benefit of mastery grading—that students had the chance to deliberately practice in areas for growth rather than just moving on after the topic was covered. This is consistent with students in generally feeling like they understood *how* to improve their grade (Item 5).

Conclusion

In mastery grading, students' attitudes towards mastery grading did seem to improve as they became more familiar with the system. Buy-in was as important as how an instructor sets up their grading system. While there was a lot of work needed to create more assignments the grading was accomplished faster. There was no discussion about partial points. Overall, the instructor considered the conversion from traditional grading a success. We now discuss some guidance for instructors who may wish to implement mastery grading in their own courses.

Challenges and Practical Implications

There were three major challenges: buy-in, course preparation, and mastery thresholds. Rojas uncovered these challenges over the semester, but being open-minded and flexible, the instructor rectified the issues.

Rojas severely underestimated the amount of buy-in required. In the second week of class the instructor received a strongly worded email that was signed by several students in the class

requesting that the class be changed to traditional grading. Some concerns that students had were that “bad test takers” were significantly penalized in the mastery grading system. Their other concern was that the system would discourage teammates from contributing to the project that were performing poorly in other non-project related categories (homework,quizzes, exam). Lastly, they voiced their concern over the limited number of tokens that students could use for revisions.

After receiving the email from the students, the instructor set aside time at the beginning of the next class session to discuss the grading scheme. The instructor walked through the reasoning of the grading scheme. For example, the exams would not penalize students because the *exam* category was computed as the maximum between the midterm and final; thus giving students two opportunities to perform on exams. Next, we built the system to incentivize students to retake assignments. Rojas would reiterate that the grading system would change if issues arose that would be unfair. For example, Rojas increased the token limit as soon as the students needed more revisions to master certain topics. The grading scheme had some issues, but over the semester things got fixed.

The course preparation required more time and effort than expected. The initial challenge was coming up with a way to map the mastery grades (NR/LP/HP) to letter grades. After reviewing several examples from Nilson [3], Rojas was able to produce Table 1. Next, was coming up with multiple versions of the same assignment or quiz. In total there were 20 assignments (including revisions for homework and quizzes). Over time, the time commitment would reduce due to the build up of a question bank. Another issue was figuring out the low pass (LP) and high pass (HP) threshold. In the beginning, the instructor set the high pass level too high. Due to revisions that issue was easily fixable by lowering the HP question difficulty in later revision assignments.

There are several considerations when transitioning a course to mastery grading. First, you must have a clear set of outcomes that are aligned with your course learning outcomes. The outcomes will be the basis for how you design your assessments. Review many examples on how to map mastery grades to letter grades. An excellent resource would be Nilson [3]. Next, you will need a plan on how to build buy-in from the students. You can set aside a few minutes each class to answer any concerns or create checklists for students to teach them how to compute their grade. Lastly, you must be prepared to change things if things don’t go as expected.

References

1. Howitz, William J., Kate J. McKnelly, and Renée D. Link. "Developing and implementing a specifications grading system in an organic chemistry laboratory course." *Journal of Chemical Education* 98.2 (2020): 385-394.
2. J. Mendez, “Standards-Based Specifications Grading in a Hybrid Course,” in *2018 ASEE Annual Conference & Exposition Proceedings*, Salt Lake City, Utah, Jun. 2018, p. 30982. doi: 10.18260/1-2--30982.
3. L. B. Nilson. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time. Stylus Publishing, LLC, 2015.*
4. L. Craugh, “Adapted Mastery Grading for Statics,” in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio, Jun. 2017, p. 27536. doi: 10.18260/1-2--27536.
5. L. Schlemer and L. Vanasupa, “Grading for Enhanced Motivation and Learning,” in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana, Jun. 2016, p. 27305. doi: 10.18260/p.27305.
6. Pascal, Jennifer, Troy J. Vogel, and Kristina Wagstrom. "Grading by Competency and Specifications: Giving better feedback and saving time." *2020 ASEE Virtual Annual Conference Content Access*. 2020.
7. Ranalli, J., & Moore, J. P. (2015). New Faculty Experiences With Mastery Grading. 2015 ASEE Annual Conference & Exposition, 26–1187.