Classification of Alternative Grading Approaches: Review and Reflections from Practice

Dr. Simon Li, University of Calgary

Dr. Simon Li is an associate professor of the Department of Mechanical and Manufacturing Engineering at the University of Calgary. His research interests include mechanical systems of buildings and engineering education.

Yves Pauchard, University of Calgary
Dr. Ahmad Ghasemloonia, University of Calgary

Dr. Ahmad Ghasemloonia received his BSc in Mechanical Engineering from the University of Tabriz in 2003 and his MSc in Applied Mechanics from Tarbiat Modares University in 2006 with a research focus on vibration condition monitoring and fault diagnosis. Between 2003 and 2009, he served as a Technical Engineer, Quality Control Engineer and Vibration Analysts at TSA, Pars Khazar, SAPCO and MAPNA Group.

Dr. Ghasemloonia received his PhD in Mechanical Engineering from Memorial University of Newfoundland, graduating with distinction in 2013 with a specialization in structural vibrations. Following the completion of his PhD, he had an Industrial Post-doctoral Fellowship at Subsurface Imaging Technology on the feasibility study of excavating salt caverns in Newfoundland.

In 2014, he joined Project neuroArm at the University of Calgary as a Post-doctoral Scholar, where he worked on biomedical engineering research projects (surgical tool design, cell vibration analysis, vibration simulation in surgical drills and surgical skill assessment using motion quality). He lectured as a Sessional Faculty at Memorial University in 2013 and at the University of Calgary in 2015 and 2016 before joining the Department of Mechanical Engineering at the University of Calgary, where he teaches courses in the areas of dynamics and vibrations.

Dr. Ghasemloonia joined Office of Dean of the Schulich School of Engineering in March 2021 as the Assistant Dean – Academic Programs and Students. In June 2022, he was appointed to the Associate Dean – Student Success and Interdisciplinary Academic Programs. In this role, he is working on students' related matters and policies. He is also the Academic Coordinator of the first year engineering program (Schulich Studio) since June 2023.

Dr. Ghasemloonia is a registered Professional Engineer (P.Eng.) in Alberta.

Classification of alternative grading approaches: review and reflections from practice

Abstract
The purpose of this paper is to review and categorize how alternative grading has been practiced in higher education and reflect on how we, as instructors in a university, apply it in their courses. In this paper, the potential issues of traditional grading are characterized in three aspects: (1) judgemental, (2) high stakes, and (3) authoritative. After conducting the literature review, we classify three categories of alternative grading practices: (1) reduction of grade granularity, (2) standardized and iterative formative assessments, and (3) democratization of grading processes, which can potentially address the issues of traditional grading. Then, we report and reflect on how alternative grading is applied in three different courses (i.e., software development, engineering dynamics, and mechanical systems in buildings). As one takeaway message of this paper, while the general principles of alternative grading are known and shared among instructors, they can imply a wide range of possibilities for testing and implementation. At the same time, despite the diverse practices, common values can emerge to support instructors for better grading practices.

Keywords: traditional grading; alternative grading; formative assessments; reflections of practitioners

1. Introduction

Grading is ubiquitous in our education system. Both students and instructors could find it difficult to imagine schools without grading. Yet, the practices of grading can lead to various challenges for student learning. To keep the discussion concise, we first want to distinguish between traditional grading (the term adopted from [1]) and alternative grading (the term adopted from [2]). In this paper, traditional grading refers to the summative assessment approaches that are commonly used to evaluate students' academic performance. Typical grading instruments include examinations and technical reports. When a grading approach is perceived as "traditional," we characterize three possible features associated with this perception.

#1: Judgemental. Traditional grading scales tend to have medium (e.g., A / B / C / D / F) to high (e.g., percentages) levels of granularity, which may not be effectively mapped to the learning outcomes. For example, it could be difficult to articulate the difference of two students who get 75% and 80%, respectively, in view of their learning outcomes. As a result, the received grades can be associated as personal values (e.g., how good am I as a student) and competition (e.g., I am better than another student). The classical study by Butler [3] has shown that the ego-involving nature of grading can disengage students from reading comments for their learning.

#2: High stakes. Assessments are high stakes if they contribute significantly to the final grades and cannot be redone. High-stakes assessments can cause student stress in learning. To cope with the high-stakes situation, students may resort to techniques that are irrelevant to learning (e.g., guessing the instructor's preferences and even cheating).

#3: Authoritative. A grading process is authoritative if the grades are completely determined by instructors or other knowledgeable persons (e.g., teaching assistants). In this situation, students can perceive that grades are less under their control. Once the grades are finalized, it will be an uphill battle for students to argue and change their grades.

While these features of traditional grading can be criticized for how they can impede student learning, we also want to balance the view that these features have their own reason in the education system. For example, summative grades are important for managing academic standings, scholarships, and graduate school admission. One-off examinations are common in professional licensure processes (e.g., engineering ethics examination). The complexity behind grading stems from the fact that grades are not just intended for student learning; they are also important for social and professional constructs [4].

In response to the negative effects of traditional grading, ungrading has been advocated to remove grades entirely from the learning process [5]. Here we interpret ungrading as a practice of refusing quantitative grading to students and focusing on qualitative feedback. Beyond ungrading as one possible option, we notice from the literature that multiple grading options can also mitigate the negative effects of grading. In this context, we refer to alternative grading as grading approaches intended to mitigate (or reduce) these characterizing features of traditional grading.

In the rest of this paper, Section 2 will present our literature review with the classification of alternative grading practices. Then, Sections 3, 4 and 5 will present our alternative grading practices in three different courses related to software development, engineering dynamics and mechanical systems in buildings, respectively. The intent of this sharing of practices is to encourage more instructor-initiated actions for alternative grading. Section 6 will summarize this paper and briefly discuss limitations and future work.

2. Literature review and classification of alternative grading

2.1. Informal collection of literature data

Data collection for the systematic review of alternative grading is not trivial. One reason is that many keywords can be used to identify relevant references (e.g., ungrading, gradeless, point-less, pass / fail). Some researchers may report improvements in grading practices aligned with our interpretation of alternative grading, but their work may not be described in the "alternative-to-the-traditional" context. Thus, we employ two less formal approaches and expect to collect

handful representative references to understand the landscape of alternative grading work without aiming to exhaust all relevant papers.

In the first approach, we used two keywords: “alternative grading” and “higher education” to search two education databases: Education Research Complete and Education Resources Information Center (ERIC). In our screening process, we selected journal papers only and skipped papers that discussed the accuracy (or statistical) improvement of grading or simply adopted new grading practices without the context of traditional grading. As a result of this search, we identify 28 references for the review. In the second approach, we utilize the search results from a recent scoping review paper on the topic of “reduced grading” [6], who identified and reviewed 23 papers. Among these 23 papers, we selected 16 that are relevant to higher education.

Some researchers have outlined various issues of grading from different aspects. Schinske and Tanner [7] have discussed the historical development and issues with grading. Kleinman et al. [8] have surveyed diverse grading practices among universities in United States. Dalziel [9] has analyzed the problems of grading from the aspect of psychological measurement. Our aim in this review is to initially identify relevant and representative papers and propose a classification of alternative grading practices, which will be overviewed in the next section. We expect that this classification can initiate some strategy for future systematic review.

2.2. Classification of alternative grading: An overview

In the literature, there can be many ideas to reduce the potential drawbacks of traditional grading. To classify these ideas, we interpret alternative grading as grading practices that are “alternative” to traditional grading. Then, we revisit the three features of traditional grading discussed earlier and propose three categories of alternative grading as follows.

Category 1: Reduction of grade granularity

High grade granularity seems to promote competition. For example, Students A and B receive 91% and 92%, respectively. Then, Student B can readily feel that “I am a better student,” but it would be hard to get this feeling if both students have the same letter grade. Also, in practice, high grade granularity tend to invite more arbitrary evaluations in grading [1]. In response, Tannock [10] has advocated to remove grades for “public university.” As the five-letter grades (i.e., A/B/C/D/F) are common, we perceive them as the lowest granularity of traditional grading. Then, any grading scales that have four or fewer levels in summative assessments can be interpreted as an alternative grading approach. One common approach is the Pass / Fail grading scheme. We can also classify the ungrading practice into this category.

Category 2: Standardized and iterative formative assessments

The distinction between formative and summative assessments may not be as clear as we perceive. For example, a quiz with a 5% weight toward the final grade may be designed as a formative assessment ,but its effect could be summative (as it statistically influences the final

grade). Maclellan [11] and Matthews and Noyes [12] have discussed different purposes and tensions between formative and summative assessments. The alternative grading practices in this category focus on the purpose of assessment for learning [13]. Examples include using standards or specifications to inform the progression of learning [2], [14] and allowing students to retake assessments [1].

Category 3: Democratization of grading processes
When instructors (or other "authoritative" evaluators) assess the student work, students are already in a passive position, waiting for grades that are less under their control. Though it is not intended for the education environment, a simple act of assigning grades may already put graders in a powerful position. By democratization, we (as instructors) basically release our "grading power" in the courses and engage students in the grading processes. Examples include contract grading, self-grading, and peer evaluations.

In our view, these categories of alternative grading can address the negative features of traditional grading discussed earlier. First, both lower grade granularity and democratization of grading can make grading less judgemental as students are not compared by hair-splitting differences, and they can have more control in grading. Second, standardized and iterative formative assessments can potentially mitigate the high-stakes nature of traditional grading because students can observe, through repeated practices, how they achieve the required standards for summative assessments. Third, democratization of grading processes is a direct response to the authoritative nature of traditional grading. It can show students how instructors "give in" their grading power and want to focus on the progression of student learning.

In the next three sections, we will review the literature based on these three categories, where we will discuss more details of alternative grading practices and their effects to student learning.

2.3. Reduction of grade granularity

In the context of grading design projects, Thompson et al. [15] from their study suggested the use of 3 or 4-point rating scales for grading consistency. To reduce study stress and nurture lifelong learning skills, gradeless and Pass / Fail systems have been applied for junior university students. However, in an introductory psychology course, Napoli and Raymond [16] showed that the ungraded condition reduced study motivation and resulted in lower academic performance. The studies by McMorran et al. [17] and McMorran and Ragupathi [18] examined the use of the Pass / Fail system for the first semester. While they found a reduction in stress and an increased willingness to try new subjects among first-year students, they also reported weaker learning attitudes and behaviors. In business education, Kjærgaard et al. [19] also showed that the Pass / Fail system for first-year students did not improve academic performance. The term "ambivalence" was used to describe the "mixed feelings" of gradeless learning [20], including:

- Less study stress now but concern about academic performance in the future,
- More intrinsic motivation with the subject but putting less effort and time for study, and
- Being part of the learning community but experiencing a loss of individual identify.

Pass / Fail or gradeless systems have also been applied in postgraduate contexts such as medical programs [21], [22] and psychology [23], where students tend to be high-achieving and have specific goals in their study. Two common observations emerge from these research studies. On the one hand, the Pass / Fail system could slightly reduce academic performance (or at least not improve it). On the other hand, the Pass / Fail system did not affect the passing rates of professional exams.

Overall, it seems less controversial to state that the Pass / Fail system can reduce student stress and improve mental wellbeing [21]. Chan and Luk [24] have also reported that qualitative grades are better than numerical grades in assessing holistic competencies of students. While gradeless learning may promote intrinsic motivation [19], it does not naturally translate into positive learning attitudes [18]. Arguably, when students are accustomed to the grading culture before entering higher education, traditional grades remain an effective means to motivate students to study diligently.

2.4. Standardized and iterative formative assessments

In this category, alternative grading approaches are applied for formative assessments or assessment for learning [13] through standardized and iterative feedback. In the literature, standardization approaches include specifications grading [14] and standards-based grading [2], which, in our view, involve an explicit mapping between learning outcomes and assessment results. Such standardization has also been commented on as an essential element for competency-based education [25]. Generally, the literature has reported positive learning effects in various course contexts such as mathematics [26], [27], chemistry laboratory courses [28], and software projects [29]. Standardization can also guide students in their learning progression [30] and facilitate positive communications between students and instructors [31].

Besides standardization, it is common to allow students repeat their work or assessments for the purpose of learning. The arrangements of such repetitions can vary significantly in the literature. For learning with intensive writing, it seems that formative feedback without grades can motivate students in their writing work [32], [33]. For technical courses with intensive subject content, one approach is to implement ungraded quizzes or assignments so that students can practice and prepare for their summative assessments [34], [35]. In a marketing course, Mittal [36] used multiple quizzes, where students were allowed to bring hand-written notes, to replace the final exam. In a math course, Cherepinsky [37] proposed a strategy by giving “right” or “wrong” only as feedback, where students with the “wrong” answer need to find solutions for iterative learning. Leggett et al. [38] reported that repeating the assessments of building-block math concepts in quizzes can lead to positive learning outcomes.

We should note that standardized and iterative formative assessments may not work for all students. It remains unclear to us if standards-based grading with low-stakes assessments can motivate students in learning intrinsically (or if students are still motivated primarily by the small

part marks for their summative assessments). Also, repetitions may promote procrastination in learning. For example, the study by Patron and Smith [39] showed that students in the "average-of-all-scores" grading scheme tended to provide better answers in the first attempts than students in the "highest-of-all-scores" (within three attempts). The study by Arnold [40] showed that students who attempted failed examinations multiple times did not necessarily demonstrate positive learning in the future. They argued that resitting for examinations could promote procrastination in learning. The study by Koh et al. [41] showed that graded (versus ungraded) assessments can motivate students better for pre-class preparation.

2.5. Democratization of grading processes

In this category, instructors delegate the "grading power" to students. One approach is self-grading, where students can grade their own work using some standards or rubrics provided by instructors [42], [43]. Popovich et al. [44] asked students to assign grades for their assignments based solely on graders' comments, and they reported a positive influence by this intervention on learning.

Some self-assessment approaches focus on meta-cognition or self-regulated learning, such as reflections on learning habits [45], [46], reflection-based self-evaluation [47], and reflections on feedback [48]. Contract grading (or learning charter) is another approach, where students can make agreements with instructors on their learning and expected achievements [49], [50]. Other ideas include peer evaluation for summative assessment [51] and student-instructor co-assessment [52].

By involving students in the grading process, certain researchers reported positive changes in students' motivation and responsibility [45], [46], [50]. Self-evaluation can be seen as a resource-saving approach for adult learners [42]. However, self-evaluation does not necessarily result in self-regulated learning [48], where grades can still be perceived as a motivating factor [47]. McClam and Sevier [53] analyzed how student-derived grades can disrupt existing power relations between instructors and students.

3. Practice and reflection I: Software development

In a 2nd-year software development course for Engineering students in the Digital Minor the focus is on object-oriented, event-based (graphical user interface), network (web server), and database programming. One approach that works well for programming courses is critique-driven assessment, discussed in Chapter 8 of [5]. In this approach, students submit work, obtain feedback, address the feedback, and resubmit until the submission is deemed complete. For software development courses, this is a natural fit because it mirrors the iterative approach to software development. The final grade is based on:

1. Progress: How far did students get (number of topics completed and difficulty of topics)?
2. Quality: How good is the work submitted? Are early errors corrected later?

3. Effort: How steadily did students submit work during the term?

In the Winter 2023 offering of this course, I piloted the following approach for assessment. The final grade was composed of traditional quizzes (40%, best 4 out of 5) and assignments (60%), with assignments being conducted using critique-driven assessment. There were 8 possible assignments on 5 coding topics, with increasing difficulty levels within the assignments. Although no deadlines were enforced, students received a suggested timeline outlining weekly submissions of either new material or addressing feedback. After the term break, students received a feedback summary similar to the final feedback summary (see below). Code was distributed using GitHub Classroom, and GitHub issues were used to track feedback for students to work on. To assess code functionality, the assignment repository contained unit tests that students could run, with passing all tests signaling satisfactory code functionality.

The work cycle for students and the instructor followed these steps:

1. Students copied GitHub Classroom link from the learning platform and accepted the assignment.
2. Students:
   2.1. Cloned the code repository to their laptop/PC.
   2.2. Worked on code, ran unit-tests, and added documentation in a markdown file.
   2.3. Committed to git and pushed to GitHub.
   2.4. Submitted GitHub repository link on the learning platform.
3. Instructor:
   3.1. Checked for new submissions on the learning platform.
   3.2. Reviewed code and documentation.
   3.3. Recorded feedback text and severity level in local database. Severity levels were numerically encoded and represented different types of feedback: problem, suggestion, neutral, and praise.
   3.4. Shared feedback on GitHub with GitHub issues and alerted students on the learning platform by releasing feedback.
4. Students reviewed feedback and went to 2-2 for improvement.

This cycle was repeated until the instructor deemed the assignment complete. While waiting for feedback, students were encouraged to start the next assignment as the focus was to submit new work weekly.

Figure 1 shows the custom database executable (implemented in Python) used to record the feedback by the instructor. Each entry captures the date/time of the submission on the learning platform, student information, assignment, and feedback. Assignment difficulty and feedback severity are linked in the database. Note that multiple feedback entries are often assigned to a single student submission.

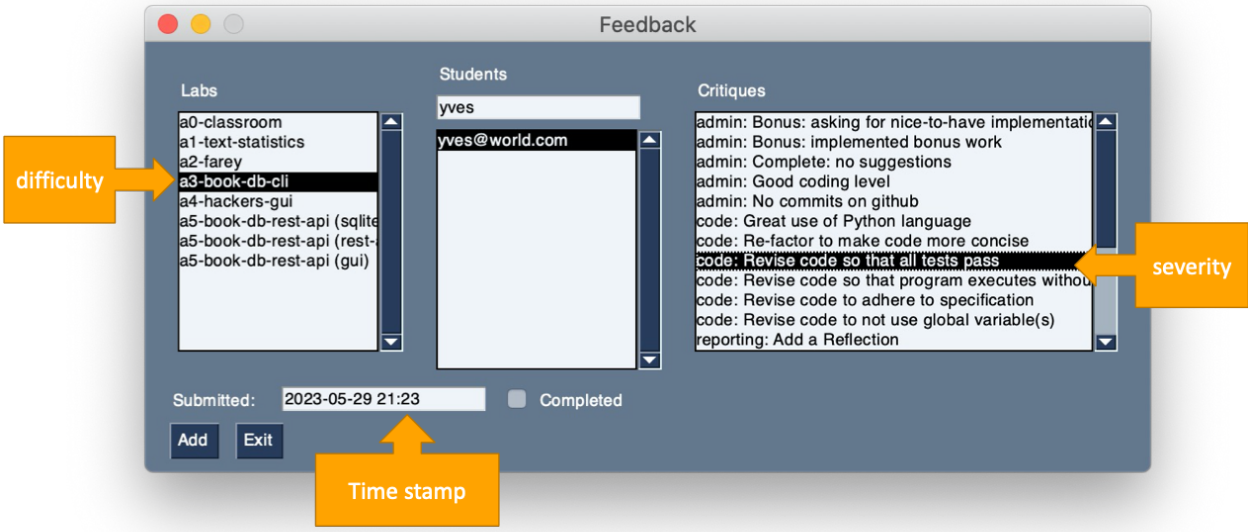

Figure 1: Desktop user interface for instructor to record student feedback

Feedback was provided to the student using GitHub issues. The feedback text in the database becomes the title of the issue. Additional information was provided, including linking of code lines and markdown-formatted hints and requests. Using GitHub tools ensured that feedback could be specific to the student's submission, clear, and leave a well-documented trail for the interaction between student and instructor. Students were instructed to reference the GitHub issue in their commits to link their improved code to the feedback.

After term break and at the end of the course, the series of feedback entries in the database were converted to a report containing: Progress, quality, and effort. Figure 2 illustrates an example.

```
--- Overall ---
score: 64
--- Progress ---
number of topics completed: 5/5
number of assignments completed: 6/8
number of bonus items: 0
--- Quality ---
number feedbacks: 15
number of praise feedbacks: 0/15
number of basic problem feedbacks: 4/15
--- Effort ---
number of submissions: 12/13
submission span: 82 days
submission variability: 7.0 days (goal: <8 days, smaller is better)
```

Figure 2. A sample of feedback to students

The score is a weighted sum of different statistics that are then converted into the final mark. An important feature of this score is that feedback severity fades over time. Hence, if basic problem feedback disappears towards the end, it does not impact the final score heavily.

The initial implementation of ungrading using critique driven assessment was successful for both students and myself. In conversations, students reported that this method aligned with their learning style and contributed to improving their programming skills. They appreciated that feedback was constructive and showed new ways to write code.

Furthermore, students enjoyed the fact that the submit-resubmit cycle made it acceptable to make mistakes without penalties. Additionally, revisiting their work and improving based on feedback was perceived to improve the quality of their work. Many struggled with the absence of enforced deadlines, mainly because other courses had strict schedules. For others, having no deadlines provided flexibility and was a blessing. Students expressed a desire for timely feedback, which is understandable in today's instant gratification world. Additionally, students expressed the need for more frequent summary reports to gauge their progress.

From an instructor's perspective, critique-driven assessment has many positive aspects. First, it creates a positive work and learning environment. Student interactions focus on course content assignment problems rather than bargaining for better grades, making the classroom feel less stressful. Second, feedback can be tailored to individual students. Some students require focus on solidifying the foundations, while others are ready to be challenged further to improve their current knowledge and skills. Like a good coach in sports, the instructor can find the current limits and help students push themselves just a bit further – this is how humans improve.

However, critique-driven assessment also presents challenges for instructors. Keeping students on track requires more effort as there are no deadlines. Since the course is part of a minor and often the 6th course for students in the term, the workload is already high. To help students get organized, a suggested schedule with weekly milestones was provided. Also, lectures start with a short logistics block to address questions about schedule and organization. Finally, weekly emails are sent out with the plan for the coming week. Another challenge is the need for a short turn-around time when providing feedback. With approximately 60 students and weekly submissions, there are many submissions to provide feedback on, which requires good organization on the instructor's side. In this initial iteration, teaching assistants were not involved, as I had to iterate and evolve the method during the term. In future course deliveries, involving teaching assistants will be useful. To streamline the feedback process further, it may be necessary to reconsider the number of tools used. Ideally, having one web-based application that serves as a central hub for all student and instructor work would simplify the process.

In conclusion, piloting critique-driven assessment was a fulfilling exercise and sparked my interest in improving this method and expanding it to other courses.

4. Practice and reflection II: Engineering dynamics

Engineering Dynamics is a challenging course undertaken by students majoring in Mechanical Engineering, Biomedical Engineering, Chemical Engineering, Geomatics Engineering, and various minors such as Aerospace and Mechatronics at our institute. Typically, this course is scheduled in the third semester, following the completion of Statics. The curriculum is divided into two main modules: Particle Dynamics, which is covered before the midterm, and Rigid Body Dynamics, which follows the midterm examination. The assessment structure includes eight quizzes, one midterm, a makeup midterm, and a final exam.

Traditionally, the course plan includes five quizzes prior to the midterm and three quizzes post-midterm. This plan allows students to study Particle Dynamics before the midterm and study Rigid Body Dynamics after the midterm. The course typically experiences a failure rate of approximately 20%, with an average grade falling within the range of 2.5 to 2.9 out of 4. In response to this challenge, one of the authors of this paper with another colleague implemented a strategy in the Fall semesters of 2019 and 2020 aimed at elevating the overall outcomes of the course. This involved offering second-chance quizzes and midterms to students who struggled with the initial assessments.

As described in a prior study authored by one of the contributors [54], the rationale behind incorporating second-chance exams extends beyond the improvement of grades for struggling students. It serves as a pedagogical tool to afford all students an additional opportunity to learn the concepts tested in these assessments which will be used as the basics in the next modules of the course. The midterm, constituting 25% of the overall course grade, has historically posed a challenge for students as the initial major assessment. Additionally, the concepts covered in the midterm pertain to particle dynamics, a foundational module for the subsequent segment of the course, rigid body dynamics. This educational philosophy aligns with Bloom's concept of mastery learning [55], emphasizing the importance of achieving a mastery level in foundational concepts before progressing to subsequent chapters that build upon prior knowledge. Research, including studies cited in references [54], [56], has consistently demonstrated higher academic outcomes in subsequent assessments and the final exam for students who engage with the makeup assessment before the final exam.

In the Fall of 2023, the course instructor implemented a refined approach to the makeup midterm, tailoring it to address the unique needs of two distinct student groups. Firstly, students who were unable to participate in the original midterm due to legitimate reasons, such as illness, were provided the opportunity to take the makeup midterm. For this group, the makeup midterm constituted 100% of their midterm grade. Secondly, for students who did not perform well on the original midterm, the makeup midterm accounted for 70% of the midterm grade, while the original midterm retained a 30% weighting. To ensure consistency in difficulty and rigor, the instructor simultaneously prepared both the makeup midterm and the original midterm. The two exams featured distinct sets of questions at the same level of difficulty. The makeup midterm was scheduled two weeks after the original midterm. Furthermore, the instructor released the grades for the original midterm a week after the examination date. This timing provided students

interested in opting for the makeup midterm with a week to review and prepare for the makeup midterm.

In the Fall 2023 semester, a total of 266 students were enrolled in the course. Thirty-four students, constituting 12.78% of the class, chose to write the makeup midterm, where five of these students did not participate in the original midterm. The overall class average in the original midterm was 70.79%. However, the 29 students who took both the original and makeup midterms displayed an average of 50.96% in the original midterm, having a 19.83% average below the class average. In a promising trend, the performance of these 29 students in the makeup midterm showcased significant improvement, with an average of 68.56%. This 17.6% enhancement from their scores in the original midterm, narrowing the gap to just 2.23% below the class average in the original midterm is notable. Among the 34 students who participated in the makeup midterm, only one student scored lower than their original midterm. Comparing these results to a pilot project conducted by one of the authors and their colleagues in Fall 2019, the current improvement of 17.6% demonstrates positive strides. Notably, in the earlier project, there was a 23% improvement, with 28.4% of students opting to write the makeup midterm.

In the Fall 2023 offering of the course, the final exam average for the entire class stood at 73.21%. However, students who opted for the makeup midterm demonstrated an average of 61.58% on the final exam, marking an 11.63% difference below the class average. A comparison can be drawn with the Fall 2019 offering, where students who undertook the makeup midterm only scored 3.6% below the class average on the final exam. This disparity between the two offerings can be attributed in part to a distinct instructional approach. The absence of makeup quizzes in the more recent offering may have contributed to the increased difference between the final exam scores of students who took the makeup midterm and the overall class average.

While the inclusion of a makeup midterm requires additional time and effort from both instructors and teaching assistants for designing, invigilating, and grading exams, the instructional team firmly believes that this approach alleviates stress for students during the original midterms. Having conducted this pilot project three times, the team recognizes the potential for further efficiency improvements.

One suggestion for enhancing the efficiency of the project involves reconsidering the grading scheme for students who opt to write the makeup midterm. Shifting away from the current 70%/30% grading allocation in favor of a more flexible approach could align better with the mastery learning goal of the project. If students invest additional time after the original midterm to grasp the course concepts, it might be beneficial to consider the better grade between their original midterm and the makeup midterm. While this adjustment has the potential to motivate a higher percentage of students to choose the makeup midterm option, it does present additional grading tasks for teaching assistants.

5. Practice and reflection III: Mechanical systems in buildings

This course is a technical elective with approximately 20-25 students. The primary objective is to introduce and prepare students for the building and HVAC (heating, ventilating and air conditioning) industry, thus emphasizing professional development. One alternative grading practice in this course is self-grading quizzes, which were piloted in Fall 2022 and Fall 2023.

Self-grading quizzes took place during the lectures (1 hour and 15 minutes), consisting of technical calculations and multiple-choice questions. Students had 45 minutes to complete the quiz and then self-graded their work within the remaining 30 minutes. I discussed each question on the spot, providing answers and grading guidance. I asked students to use different colored pens to distinguish their writing during self-grading. Afterward, I collected all quizzes for grade recording, while also double-checking the fairness of self-graded grades. If necessary, I would adjust the grades and communicate the reasons with the students.

One positive observation is that the self-grading period (30 minutes) appears to be the most effective time for student learning in the semester. Students showed eagerness to know the answers and checking their own work. For instance, in the final exam, repeating a similar question that was previously done weakly in the quiz resulted in improved performance by students. Additionally, this practice helps save grading time.

An additional, less obvious benefit is that students with low confidence seem to benefit from this process. Rather than being informed by the instructor about their errors, students could see and reflect on their mistakes. These students often marked themselves lower than expected, where I needed to mark up their grades. They also seemed more willing to participate in class discussions. However, further observation is needed to determine if this effect on students with low confidence is consistent or occasional.

In terms of academic performance, I did not observe significant decline or improvement. It is expected that self-awareness from self-grading may not necessarily lead to intrinsic motivation or better learning habits. However, this self-grading practice did not result in grade inflation or cheating. Questions in the quizzes were mostly calculation and concept-based, where answer correctness was clear. The low-stakes nature of the quizzes and the “grading-together” environment likely reduced the incentive of cheating.

One negative observation is that self-grading quizzes comprised a significant portion of the final grades in the end (Fall 2022: two quizzes totaling 20% of the final grade; Fall 2023: three quizzes totaling 24% of the final grade). This may have led some students to take the final exams less seriously. This aligns with literature suggesting that formative and summative assessments should be more distinctly designed and arranged. While the intention was to make self-grading quizzes formative, their overall summative effect on the final grade (20%-24%) may have adversely affected student learning.

6. Closing remarks

This paper has classified alternative grading practices into three categories: (1) reduction of grade granularity, (2) standardized and iterative formative assessments, and (3) democratization of grading processes. We hope that this categorization can provide a reasonable spectrum for instructors to consider and evaluate different grading practices. As shared by Feldman [1], grading can be strongly tied to the personal values and preferences of instructors. We hope that this categorization and our sharing of practices can promote more bottom-up actions and support instructors in identifying the rationale for their grading practices in response to the issues of traditional grading.

This paper has two limitations. First, the literature review process is rather informal compared to typical review practices. With this paper as an initial step, we can identify various scopes of alternative grading, which can support a more systematic review in the future. Second, the paper does not convey empirical data of our grading practices. We should note that engineering education is not a common context in the literature on alternative grading or ungrading, unlike first-year programs, math education, and writing. Further research studies would be desirable to investigate specific issues of alternative grading.

References

[1] J. Feldman, *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*, 1st ed. Thousand Oaks: Corwin Press, 2018.

[2] D. Clark and R. Talbert, *Grading for Growth: A Guide to Alternative Grading Practices That Promote Authentic Learning and Student Engagement in Higher Education*, 1st ed., vol. 1. London: Routledge, 2023. doi: 10.4324/9781003445043.

[3] R. Butler, "Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance," *J. Educ. Psychol.*, vol. 79, no. 4, pp. 474–482, 1987, doi: 10.1037/0022-0663.79.4.474.

[4] R. Lynch and J. Hennessy, "Learning to earn? The role of performance grades in higher education," *Stud. High. Educ.*, vol. 42, no. 9, pp. 1750–1763, Sep. 2017, doi: 10.1080/03075079.2015.1124850.

[5] S. D. Blum, A. Kohn, and T. Saffel, *Ungrading: why rating students undermines learning (and what to do instead)*, First edition. in Teaching and learning in higher education. Morgantown: West Virginia University Press, 2020.

[6] D.-A. Normann, L. V. Sandvik, and H. Fjørtoft, "Reduced grading in assessment: A scoping review," *Teach. Teach. Educ.*, vol. 135, p. 104336, Dec. 2023, doi: 10.1016/j.tate.2023.104336.

[7] J. Schinske and K. Tanner, "Teaching More by Grading Less (or Differently)," *CBE—Life Sci. Educ.*, vol. 13, no. 2, pp. 159–166, Jun. 2014, doi: 10.1187/cbe.cbe-14-03-0054.

[8] S. B. Kleinman, M. B. Leidman, and A. J. Longcore, "The changing landscape of grading systems in US higher education," *Perspect. Policy Pract. High. Educ.*, vol. 22, no. 1, pp. 26–33, Jan. 2018, doi: 10.1080/13603108.2017.1279692.

[9] J. Dalziel, "Using Marks To Assess Student Performance: Some Problems and Alternatives," *Assess. Eval. High. Educ.*, vol. 23, no. 4, pp. 351–66, 1998.

[10] S. Tannock, "No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university," *Stud. High. Educ.*, vol. 42, no. 8, pp. 1345–1357, Aug. 2017, doi: 10.1080/03075079.2015.1092131.

[11] E. Maclellan, “How convincing is alternative assessment for use in higher education?,” *Assess. Eval. High. Educ.*, vol. 29, no. 3, pp. 311–321, Jun. 2004, doi: 10.1080/0260293042000188267.
[12] R. Matthews and A. Noyes, “To grade or not to grade: balancing formative and summative assessment in post-16 teacher trainee observations,” *J. Furth. High. Educ.*, vol. 40, no. 2, pp. 247–261, Mar. 2016, doi: 10.1080/0309877X.2014.953456.
[13] R. E. Bennett, “Formative assessment: a critical review,” *Assess. Educ. Princ. Policy Pract.*, vol. 18, no. 1, pp. 5–25, Feb. 2011, doi: 10.1080/0969594X.2010.513678.
[14] L. B. Nilson, *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*, First edition. Sterling, Virginia: Stylus, 2015.
[15] M. K. Thompson, L. H. Clemmensen, and B. Ahn, “The effect of rubric rating scale on the evaluation of engineering design projects,” *Int. J. Eng. Educ.*, vol. 29, no. 6, pp. 1490–1502, 2013.
[16] A. R. Napoli and L. A. Raymond, “How Reliable Are Our Assessment Data?: A Comparison of the Reliability of Data Produced in Graded and Un-Graded Conditions,” *Res. High. Educ.*, vol. 45, no. 8, pp. 921–929, Dec. 2004, doi: 10.1007/s11162-004-5954-y.
[17] C. McMorran, K. Ragupathi, and S. Luo, “Assessment and learning without grades? Motivations and concerns with implementing gradeless learning in higher education,” *Assess. Eval. High. Educ.*, vol. 42, no. 3, pp. 361–377, Apr. 2017, doi: 10.1080/02602938.2015.1114584.
[18] C. McMorran and K. Ragupathi, “The promise and pitfalls of gradeless learning: responses to an alternative approach to grading,” *J. Furth. High. Educ.*, vol. 44, no. 7, pp. 925–938, Aug. 2020, doi: 10.1080/0309877X.2019.1619073.
[19] A. Kjærgaard, J. Buhl-Wiggers, and E. N. Mikkelsen, “Does gradeless learning affect students’ academic performance? A study of effects over time,” *Stud. High. Educ.*, pp. 1–15, Jul. 2023, doi: 10.1080/03075079.2023.2233007.
[20] A. Kjærgaard, E. N. Mikkelsen, and J. Buhl-Wiggers, “The gradeless paradox: Emancipatory promises but ambivalent effects of gradeless learning in business and management education,” *Manag. Learn.*, vol. 54, no. 4, pp. 556–575, Sep. 2023, doi: 10.1177/13505076221101146.
[21] R. A. Bloodgood, J. G. Short, J. M. Jackson, and J. R. Martindale, “A Change to Pass/Fail Grading in the First Two Years at One Medical School Results in Improved Psychological Well-Being,” *Acad. Med.*, vol. 84, no. 5, p. 655, May 2009, doi: 10.1097/ACM.0b013e31819f6d78.
[22] C. B. White and J. C. Fantone, “Pass–fail grading: laying the foundation for self-regulated learning,” *Adv. Health Sci. Educ.*, vol. 15, no. 4, pp. 469–477, Oct. 2010, doi: 10.1007/s10459-009-9211-1.
[23] R. M. Roberts and D. S. Dorstyn, “Does changing grade-based marking to pass-fail marking change student achievement in a postgraduate psychology course?,” *Train. Educ. Prof. Psychol.*, vol. 11, no. 1, pp. 57–60, 2017, doi: 10.1037/tep0000129.
[24] C. K. Y. Chan and L. Y. Y. Luk, “Going ‘grade-free’? – Teachers’ and students’ perceived value and grading preferences for holistic competency assessment,” *High. Educ. Res. Dev.*, vol. 41, no. 3, pp. 647–664, Apr. 2022, doi: 10.1080/07294360.2021.1877628.
[25] M. Townsley and D. Schmid, “Alternative grading practices: An entry point for faculty in competency-based education,” *J. Competency-Based Educ.*, vol. 5, no. 3, p. e01219, 2020, doi: 10.1002/cbe2.1219.

[26] J. K. Zimmerman, “Implementing Standards-Based Grading in Large Courses Across Multiple Sections,” *PRIMUS*, vol. 30, no. 8–10, pp. 1040–1053, Nov. 2020, doi: 10.1080/10511970.2020.1733149.
[27] D. Lewis, “Impacts of Standards-Based Grading on Students’ Mindset and Test Anxiety,” *J. Scholarsh. Teach. Learn.*, vol. 22, no. 2, pp. 67–77, Jun. 2022.
[28] W. J. Howitz, K. J. McKnelly, and R. D. Link, “Developing and Implementing a Specifications Grading System in an Organic Chemistry Laboratory Course,” *J. Chem. Educ.*, vol. 98, no. 2, pp. 385–394, Feb. 2021, doi: 10.1021/acs.jchemed.0c00450.
[29] R. Quintana and C. Quintana, “When classroom interactions have to go online: the move to specifications grading in a project-based design course,” *Inf. Learn. Sci.*, vol. 121, no. 7/8, pp. 525–532, Jan. 2020, doi: 10.1108/ILS-04-2020-0119.
[30] P. Grainger and K. Weir, “An alternative grading tool for enhancing assessment practice and quality assurance in higher education,” *Innov. Educ. Teach. Int.*, vol. 53, no. 1, pp. 73–83, Jan. 2016, doi: 10.1080/14703297.2015.1022200.
[31] M. H. Scarlett, “‘Why Did I Get a C?’: Communicating Student Performance Using Standards-Based Grading,” *InSight J. Sch. Teach.*, vol. 13, pp. 59–75, 2018.
[32] A. A. Lipnevich and J. K. Smith, “‘I really need feedback to learn:’ students’ perspectives on the effectiveness of the differential feedback messages,” *Educ. Assess. Eval. Account.*, vol. 21, no. 4, pp. 347–367, Nov. 2009, doi: 10.1007/s11092-009-9082-2.
[33] K. M. Armstrong, “Fluency, Accuracy, and Complexity in Graded and Ungraded Writing,” *Foreign Lang. Ann.*, vol. 43, no. 4, pp. 690–702, 2010, doi: 10.1111/j.1944-9720.2010.01109.x.
[34] E. Kitchen, S. H. King, D. F. Robison, R. R. Sudweeks, W. S. Bradshaw, and J. D. Bell, “Rethinking Exams and Letter Grades: How Much Can Teachers Delegate to Students?,” *CBE—Life Sci. Educ.*, vol. 5, no. 3, pp. 270–280, Sep. 2006, doi: 10.1187/cbe.05-11-0123.
[35] M. M. Khanna, “Ungraded Pop Quizzes: Test-Enhanced Learning Without All the Anxiety,” *Teach. Psychol.*, vol. 42, no. 2, pp. 174–178, Apr. 2015, doi: 10.1177/0098628315573144.
[36] S. Mittal, “The Tyranny of Tests: Doing Away with Tests Without Selling Your Soul to Grading Hell,” *Mark. Educ. Rev.*, vol. 29, no. 2, pp. 136–141, Apr. 2019, doi: 10.1080/10528008.2019.1610330.
[37] V. Cherepinsky, “Self-Reflective Grading: Getting Students to Learn from their Mistakes,” *PRIMUS*, vol. 21, no. 3, pp. 294–301, Mar. 2011, doi: 10.1080/10511970903147861.
[38] J. Leggett, B. Morgan, and K. K. Tang, “Repeated quizzing of basic mathematics concepts to improve grades in economics classes,” *J. Econ. Educ.*, vol. 53, no. 4, pp. 296–306, Oct. 2022, doi: 10.1080/00220485.2022.2111386.
[39] H. Patron and W. J. Smith, “Implications of Fail-Forward in an Online Environment under Alternative Grading Schemes,” *J. Educ. Online*, vol. 8, no. 1, Jan. 2011, Accessed: Jan. 12, 2024. [Online]. Available: https://eric.ed.gov/?id=EJ917869
[40] I. Arnold, “Resitting or compensating a failed examination: does it affect subsequent results?,” *Assess. Eval. High. Educ.*, vol. 42, no. 7, pp. 1103–1117, Oct. 2017, doi: 10.1080/02602938.2016.1233520.
[41] Y. Y. J. Koh, J. I. Rotgans, P. Rajalingam, P. Gagnon, N. Low-Beer, and H. G. Schmidt, “Effects of graded versus ungraded individual readiness assurance scores in team-based learning: a quasi-experimental study,” *Adv. Health Sci. Educ.*, vol. 24, no. 3, pp. 477–488, Aug. 2019, doi: 10.1007/s10459-019-09878-5.

[42] G. Stebbins, “Planning and Delivering Instruction with Increasing Class Sizes in Educational Administration Program Coursework: Modeling Leadership Skills for New Professors Transitioning from K-12 Administration,” *Educ. Leadersh. Adm. Teach. Program Dev.*, vol. 21, pp. 46–58, 2009.
[43] P. Grainger *et al.*, “Improving the quality of assessment by using a community of practice to explore the optimal construction of assessment rubrics,” *Reflective Pract.*, vol. 18, no. 3, pp. 410–422, May 2017, doi: 10.1080/14623943.2017.1295931.
[44] N. G. Popovich, K. S. Plake, and M. K. Scott, “Evaluation of Student Attitude Toward Learning From Homework Assignments in a Nonprescription Drugs Course,” *Am. J. Pharm. Educ.*, vol. 60, no. 3, pp. 275–280, 1996.
[45] B. Strong, M. Davis, and V. Hawks, “Self-Grading in Large General Education Classes: A Case Study,” *Coll. Teach.*, vol. 52, no. 2, pp. 52–57, 2004.
[46] D. Guberman, “Student Perceptions of an Online Ungraded Course,” *Teach. Learn. Inq.*, vol. 9, no. 1, Art. no. 1, Mar. 2021, doi: 10.20343/teachlearninqu.9.1.8.
[47] T. Gorichanaz, “‘It made me feel like it was okay to be wrong’: Student experiences with ungrading,” *Act. Learn. High. Educ.*, p. 14697874221093640, May 2022, doi: 10.1177/14697874221093640.
[48] M. Jackson and L. Marks, “Improving the effectiveness of feedback by use of assessed reflections and withholding of grades,” *Assess. Eval. High. Educ.*, vol. 41, no. 4, pp. 532–547, May 2016, doi: 10.1080/02602938.2015.1030588.
[49] G. Potts, “A Simple Alternative to Grading,” *Inquiry*, vol. 15, no. 1, pp. 29–42, 2010.
[50] K. Meinking and E. E. Hall, “Letting Go of Grades: Creating an Environment of Autonomy and a Focus on Learning for High Achieving Students,” *Teach. Learn. Inq.*, vol. 10, Jan. 2022.
[51] A. ArchMiller, J. Fieberg, J. D. Walker, and N. Holm, “Group peer assessment for summative evaluation in a graduate-level statistics course for ecologists,” *Assess. Eval. High. Educ.*, vol. 42, no. 8, pp. 1208–1220, Nov. 2017, doi: 10.1080/02602938.2016.1243219.
[52] M. A. Gómez-Ruiz and V. Quesada-Serra, “An Analysis of Shared Grading in Co-Assessment Practices by Teachers and Students.,” *Electron. J. Educ. Res. Assess. Eval.*, vol. 26, no. 1, 2020, doi: 10.7203/relieve.26.1.16567.
[53] S. McClam and B. Sevier, “Troubles with grades, grading, and change: Learning from adventures in alternative assessment practices in teacher education,” *Teach. Teach. Educ.*, vol. 26, no. 7, pp. 1460–1470, Oct. 2010, doi: 10.1016/j.tate.2010.06.002.
[54] A. Ghasemloonia and M. N. Singh, “Using assessments to improve student outcomes in engineering dynamics,” in *2020 ASEE Virtual Annual Conference Content Access*, 2020. Accessed: Jan. 30, 2024. [Online]. Available: https://peer.asee.org/using-assessments-to-improve-student-outcomes-in-engineering-dynamics
[55] B. S. Bloom, “Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1.,” *Eval. Comment*, vol. 1, no. 2, p. n2, 1968.
[56] J. W. Morphew, M. Silva, G. Herman, and M. West, “Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering,” *Appl. Cogn. Psychol.*, vol. 34, no. 1, pp. 168–181, Jan. 2020, doi: 10.1002/acp.3605.