Improving Student Motivation by Ungrading

Scott Spurlock
sspurlock@elon.edu
Elon University
Elon, NC, USA

ABSTRACT

Recent interest in alternative grading strategies has been increasing in the Computer Science Education community. The umbrella term *ungrading* has been used to refer to a variety of practices that de-emphasize numeric grades. In this paper we present the results of implementing an ungrading scheme that eliminates numeric grades, allows resubmission of assignments, and encourages student input into their final assigned letter grade. We administered surveys measuring student attitudes and motivation at the start and end of three different upper level Computer Science elective courses using the new grading scheme and found a significant increase in students' feelings of intrinsic goal orientation (valuing coursework for its own sake), self-efficacy (feeling able to be successful), and control of learning (taking responsibility for their own learning). We observed that, given the opportunity, most students chose to redo only a small number of assignments, and most students requested final grades within a half-letter of the instructor's estimate. Overall, compared with prior iterations of the courses that were graded traditionally, the final grade point average did not significantly increase, while students' reported level of effort did significantly increase. Comments on post-course surveys indicate that students liked the new grading scheme, and they reported improved learning and reduced anxiety.

CCS CONCEPTS

• **Social and professional topics → Student assessment**.

KEYWORDS

computer science education, assessment, pedagogy, ungrading

ACM Reference Format:
Scott Spurlock. 2023. Improving Student Motivation by Ungrading. In *Proceedings of the 54th ACM Technical Symposium on Computing Science Education V. 1 (SIGCSE 2023), March 15–18, 2023, Toronto, ON, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3545945.3569747

1 INTRODUCTION

Criticisms of grading are as old as grading itself [5]. In 1933, Crooks et al. found "much divergence of opinion on the reliability of marks, their purposes, the methods of their presentation, and even their necessity" [9]. More recent studies confirm that educational assessment can vary dramatically across different assessors and even over time by the same assessor [17]. Grades often lack validity, reliability, and objectivity [14]. In general, there is much debate in interpreting exactly what is measured by numeric grades [28].

In addition, student *intrinsic motivation* (desire to learn for its own sake) is adversely affected by grading [14]. In fact, prior work has shown that students given feedback in the form of comments only, rather than numeric scores or numeric scores in combination with comments, showed increased interest and performance on related tasks [7]. Further, the *extrinsic motivation* (focus on external rewards and punishments) provided by grades increases student anxiety and motivation to cheat [19].

Traditional numeric grading also appears to negatively affect equity, particularly among traditionally underrepresented groups. Some work has shown that grades often measure factors more closely tied to socioeconomic status and other influences than student learning [28]. Alternative grading practices can help level the playing field [11].

In answer to these issues with traditional grading, *ungrading* [4] encompasses a variety of ideas to de-emphasize grades. Our ungrading approach, which also pulls from the related and overlapping *specifications grading* [21] and *grading for equity* [11], adopts three key ideas from these strategies: (1) eliminating numeric grades, (2) allowing assignment resubmission, and (3) encouraging student input to their final assigned letter grade. (We review several alternative grading strategies in Section 2.)

In this work, we describe these grading changes in more detail and report results from implementing our ungrading approach in three upper-level Computer Science courses. We present data from pre- and post-term surveys measuring changes in student attitudes and motivations.

The following research question frames our analysis of the effects of this alternative grading scheme:

RQ: How will our ungrading approach affect student learning attitudes and motivation?

Our main contributions are (1) describing a concrete implementation in upper-level Computer Science courses of several alternative grading practices; and (2) evaluating their impact by analyzing pre- and post-term survey data gathered from students taking these courses.

2 RELATED WORK

Interest in alternative grading strategies for computer science is growing, with more publications in recent years, and a variety of workshops and panels arising to meet a perceived need for change, e.g., [6, 12, 16, 32]. The factors driving interest in new grading

SIGCSE 2023, March 15–18, 2023, Toronto, ON, Canada

ACM ISBN 978-1-4503-9431-4/23/03…$15.00
https://doi.org/10.1145/3545945.3569747

schemes include a desire for better student learning, commitment to greater equity for all students, and frustration with the time-consuming tedium conventional grading imposes on instructors. To address these issues, numerous alternative grading approaches have been proposed, under a number of different names and with some degree of overlap among the assorted labels. Mastery Grading [3] has been used in classrooms for decades and seems to be undergoing renewed interest in recent years [22, 29, 31]. The approach typically allows students more control over their learning pace, incorporates pass/fail grading, and encourages retakes of material. Contract grading has also seen decades of use [30]. Although there have been a variety of implementations, the key idea is that, at the beginning of the course, a clear set of criteria are identified that will lead to each possible end-of-term grade ("criterion-referenced measurement") [13].

For our own approach, we have been most influenced by the following three grading systems.

Specifications Grading. Specifications ("specs") grading advocates linking final course grades to the successful completion of specific learning outcomes; assessing these outcomes as pass or fail; and allowing a limited number of second chances [21]. Final course grades may be based on *bundles* of assignments, similar to a contract grading scheme that specifies specific targets for different grade levels. Multiple researchers report implementing specs grading in a variety of Computer Science courses with positive results [15, 26]. Others have adapted specifications grading for a writing-intensive course [18], combined it with mastery grading in a discrete math course [31], and created a hybrid of specifications and traditional grading applied to several upper-level Computer Science courses [27]. Berns adopts a version of specifications grading with more liberal retake policies across a variety of courses [1].

Grading for Equity. Feldman's *Grading for Equity* provides an in-depth look at a collection of practices that promote equity, including removing late penalties, using a compressed grading scale, and allowing second chances on assignments [11]. Some recent work has found that incorporating the principles of equity in grading (and course design) has improved the teaching experience as well as student motivation [2]. Allowing retakes not only supports equity, it facilitates students developing mastery, although because of difficulties scaling retakes to larger numbers of students, automated support tools are very helpful [12]. Other work finds that, given the opportunity for regrading in an introductory computer science course, while some student resubmissions were motivated purely by grade improvement, many students sought to complete or improve their work [23].

Ungrading. Recently the term *ungrading* has been used to describe a focus on eliminating grades altogether, or at least de-emphasizing their importance. Blum collects several examples of courses that have experimented with a variety of ungrading ideas [4]. Many of the chapters in the collection describe replacing grades with text feedback, encouraging students to revise assignments, and giving students input to their final course letter grade. Riesbeck adopts a strategy of providing only feedback (no grades) for programming assignments, which students incorporate in resubmissions in a continuous "do-review-redo" cycle; students receive only a final course grade based on the portfolio of critiques they have received during the semester [25]. Chu advocates a standards-based grading approach, where students receive feedback instead of scores on assignments, they are allowed to retake portions of tests they have not yet mastered, and final grades are determined collaboratively between student and instructor [8]. The focus on feedback is supported by many studies, including one recent paper reviewing the importance of the type of feedback students receive [20].

3 ALTERNATIVE GRADING SCHEME

In attempting to put many of these ideas into practice, we have adopted a modified grading scheme with the following changes from our former traditional system: eliminating numeric grades, allowing assignment resubmission, and encouraging student input to their final assigned letter grade.

In previous years, we followed a traditional model: each assignment was graded with a numeric score out the possible points, the numeric scores were averaged for each student to calculate a final percentage, which was thresholded to produce a final letter grade. For the most recent, ungraded course offerings, each assignment was instead given feedback and a label: complete or incomplete, as advocated by specs grading [21]. In practice, "complete" corresponds loosely to what would have been $\geq 80\%$ in a traditionally graded scheme. For larger assignments, we used a four-level scale: complete, nearly complete, somewhat complete, and incomplete. These levels would correspond to approximate score cut-offs of 90%, 80%, 70%, and < 70% in a traditional system.

Course work falls into four categories: daily "lab" activities (small, in-class programming assignments primarily marked for completion), homework assignments (larger, out-of-class programming projects), quizzes (in-class, auto-graded assessments), and a final project. To provide clarity to students and reduce anxiety about a novel grading scheme, we took from contract grading the idea of clearly pre-defining a set of expectations for students to achieve a given final grade level, with no curves or percentages [30]. In the syllabus, students are provided a breakout of goals for these assignments and how these goals correlate to final course letter grade targets (A, B, or C). For example, if a student is targeting a final grade of B, they should complete all but three of the lab activities; completely meet requirements for two and nearly meet requirements for three of the homework assignments; score at least 80% correct on four of five quizzes; and nearly meet requirements for the final project. Table 1 shows the details of the mapping of assignment goals to final letter grade targets.

Redo Requests. Table 1 also provides guidelines on how many "redo requests" students can use to improve their work and meet assignment goals. For example, at the A level, students may revise and resubmit three labs, one homework, and two quizzes. Philosophically, we are drawn to the idea of allowing students unlimited attempts to revise work; however, concerns about an overwhelming number of requests led us to limit this option. In Section 5.2, we will discuss the number and distribution of redo requests.

Final Grade Determination. One idea advocated by many instructors as part of the ungrading movement is that students should be able to help determine their own final grades (if a final grade is required

Assignment Type (qty)	Goal	A Level	B Level	C Level
Lab Activities (25)	Complete	All but 2	All but 3	All but 4
Lab Redo Requests	Use no more than	3	5	7
Homework Assignments (5)	Meets requirements	Completely for all	Completely for 2 Nearly for 3	Nearly for 3 Somewhat for 2
Homework Redo Requests	Use no more than	1	2	3
Quizzes (5)	Complete (≥ 80%)	All	All but 1	All but 2
Quiz Redo Requests	Use no more than	2	3	4
Final project (1)	Meets requirements	Completely	Nearly	Somewhat

Table 1: Each type of assignment is associated with a goal that can be met at different levels corresponding to the level of accomplishment for each final grade level. Students are encouraged to use redo requests to meet an assignment goal that may not have been met on the initial attempt.

by the institution) [4]. In our grading scheme, while the grade targets are clearly defined in Table 1, there is room for discussion in cases where a student has met some of the criteria for one grade target and some of the criteria for another grade target. Students were encouraged to have a voice (via survey and, if needed, in-person discussion) in the determination of their final course grade by making a case for what it should be, though with the caveat that the instructor reserved the right to make the final decision. In Section 5.3, we will discuss how student-requested grades generally matched up with instructor assigned grades, as well as how student feelings of control over their own learning increased.

4 SURVEY DEVELOPMENT

To gather data about student learning attitudes, we adapted the Motivated Strategies for Learning Questionnaire (MSLQ), which consists of questions scored on a 7-point Likert scale [24]. The questions are designed to measure several categories (scales) of student motivation and learning approaches. The developers of the MSLQ note that each scale is modular by design so that particular studies can focus narrowly on the topics of interest [24]. Other Computer Science Education research has also incorporated a subset of MSLQ categories to measure changes in student attitudes [10]. For our study, we included six scales specifically focused on student motivation (31 questions total):

(1) Self-efficacy for Learning and Performance (8 questions): how confident is a student in their abilities to learn the course material and to be successful and perform well in the course?
(2) Task Value (6 questions): how valuable does a student find the subject matter?
(3) Test Anxiety (5 questions): how much do negative thoughts and emotions affect students while taking tests?
(4) Intrinsic Goal Orientation (4 questions): to what extent is a student motivated by course tasks themselves as opposed to external factors.
(5) Extrinsic Goal Orientation (4 questions): to what extent is a student motivated by external factors such as grades or competition?
(6) Control of Learning Beliefs (4 questions): how much does a student feel that their own efforts will enable them to learn the course material?

In addition to these MSLQ-based questions, students were also asked as part of the post-survey several questions specifically about their experience with the new grading scheme and an open-ended free-text response question soliciting their thoughts about the course. The following section describes the details of the collected survey data.

5 RESULTS

We examined three upper-level Computer Science elective courses during the 2021-2022 academic year at a liberal arts institution in the southeastern United States. The courses included Web Development (Fall 2021, 27 students), Data Mining and Machine Learning (Spring 2022, 47 students across two sections), and Computer Vision (Spring 2022, 23 students). Classes met for 100 minutes twice per week. Combined there were 97 students, primarily in their 3rd and 4th years, with a mix of Computer Science majors, Computer Science minors, and Data Science minors enrolled in the courses. The courses employed an active learning approach, with class time split between short lectures, interactive code demonstrations, and daily hands-on coding activities.

To analyze changes in student learning attitudes, we administered a pre- and post-survey during the first and last weeks of the semester to students in these alternatively graded courses. We retained the data for 84 students who took both.

5.1 Student Motivation Changes

For the survey questions measuring student motivation, average scores on five of the six scales increased from the start to the end of the course and decreased for the measure related to test anxiety. Figure 1 shows the average change in each category. Of these changes, three categories showed statistically significant differences using a dependent t-test for paired samples: Intrinsic Goal Orientation ($p=0.001$), Self-efficacy for Learning and Performance ($p=0.001$), and Control of Learning Beliefs ($p=0.01$).

Intrinsic Goal Orientation. The category with the largest change is related to students' intrinsic motivation. Of the four questions in the scale, the one with the largest increase was "In a class like this,

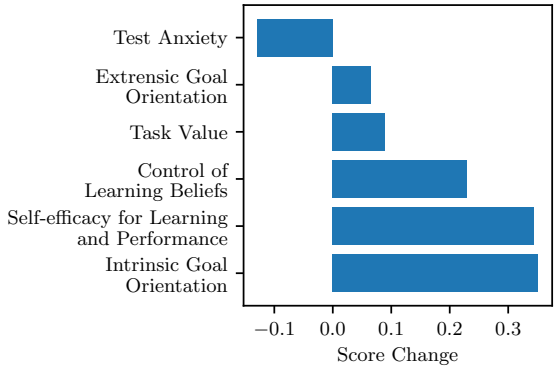

Figure 1: Average change for each measure of student motivation from pre- to post-questionnaire. The largest changes (bottom three categories) were statistically significant ($p \leq 0.01$).

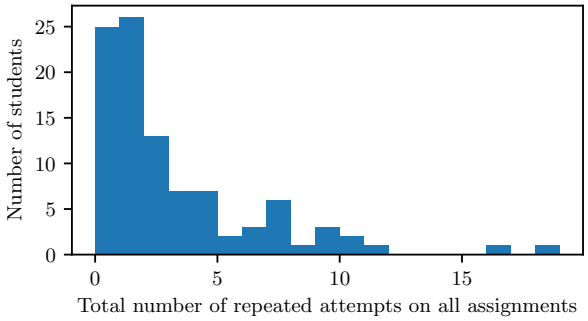

Figure 2: The distribution of the number of total redo requests on assignments. 52% of students used either zero or one redo, and 80% used fewer than five.

I prefer course material that really challenges me so I can learn new things." We believe that this increased appetite for challenge is tied to students receiving text feedback only, rather than numeric scores, on their assignments. This finding echoes prior research that has found that decreasing the focus on grades can lead to improved intrinsic motivation [7].

Self-efficacy for Learning and Performance. The second largest change relates to increased confidence of being capable of learning the material. Within this category, the question with the largest increase was "I'm confident I can understand the most complex material presented by the instructor in this course." We believe that providing opportunities to redo assignments was especially helpful in driving this improvement by allowing students to learn from initial mistakes.

Control of Learning Beliefs. The third largest change measures the extent to which students linked their own efforts to outcomes. The question with the largest average increase within the scale was "If I study in appropriate ways, then I will be able to learn the material in this course." Of note, prior work has observed a statistically significant *decrease* in Control of Learning Beliefs in traditionally graded introductory Computer Science courses, possibly attributable to students coming to doubt their own ability to achieve success [10]. Student comments suggest that the availability of redo requests, as well as the well-defined grade targets (Table 1) and ability to have a voice in their final grade contributed to this increase.

Other Categories. The changes in the other three categories were not statistically significant. The Task Value category was the highest scored scale overall on the pre-survey (suggesting that students taking upper-level Computer Science electives already see the value in the courses), so while the average increased on the post-survey, there was not much additional room at the top of the range. Extrinsic Goal Orientation reassuringly showed a negligible increase, while Test Anxiety was the only category to drop, though not enough to be significant. Possibly a stronger effect on anxiety could have been observed with differently worded survey questions: the MSLQ questions focus narrowly on *test* anxiety. On the pre-survey, this scale averaged the lowest score overall, indicating that fewer students are directly struggling with this type of anxiety. A more general (non-MSLQ) question on the post-survey ("I felt less anxiety about this course because of the grading approach") received an average score of 6.08 out of 7, where 7 indicates "strongly agree."

5.2 Second Chances

Students were given the option to resubmit work as shown in Table 1. While we were initially concerned about being overwhelmed with regrading, counting up all redo requests for all assignment types (labs, homeworks, etc.) for each student in a given course, 52% of students used either zero or one redo total, and 80% used fewer than five. Figure 2 shows the distribution of students according to how many total repeated attempts they made on assignments in one course. Most commonly these "less than five" students used a redo to retake a quiz that did not go as well as they hoped, often after coming to office hours to talk about a confusing topic. Also common were students who did not completely meet the requirements on a homework assignment on their initial attempt; after reading the feedback (and sometimes coming to office hours), these students were generally able to resubmit and completely meet the requirements.

A few students used many redo requests. These students had a similar profile: they tended to struggle with deadlines and class attendance. The large number of second attempts for these students primarily represents late first submissions, which were originally marked incomplete. These are students who would have received zeros on these assignments with a traditional grading scheme, leading, in all probability, to a failing grade for the semester. Under the proposed grading scheme, these students were able to complete the work and receive passing course grades.

5.3 Final Grades

Although the usefulness and exact meaning of grades can be debated [4], our institution, like most, requires each student to be assigned a final letter grade for each course. As described in Section 3, we arrived at this final grade with reference to Table 1 and incorporated each student's input. Our preference would be to simply

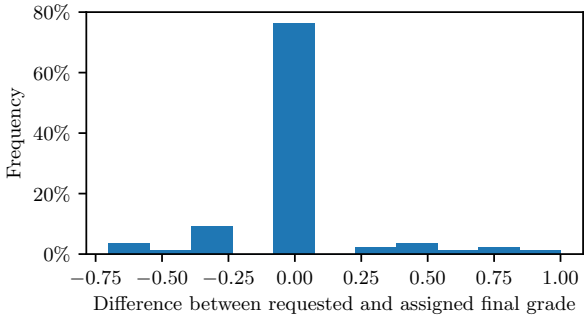

Figure 3: The distribution of differences between each student's requested final course grade and the grade they were actually assigned (converted to a 4-point scale). Negative values indicate a student received a higher final grade than requested.

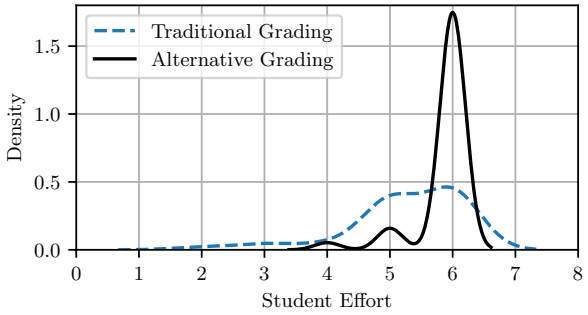

Figure 4: The kernel density estimate of students' self reported course effort (0 – 6 scale) for traditionally graded and alternatively graded courses. Compared with prior offerings of the same courses by the same instructor, students reported more consistent effort under the new grading scheme.

assign each student whatever grade they indicated; however, we made adjustments when a student's grade justification did not conform well to Table 1. Interestingly, this adjustment worked in both directions, with students under- and overestimating their accomplishments for the semester. Most commonly though, our estimate matched the student estimate, or the student's justification was convincing enough to sway the outcome.

Figure 3 shows the distribution of differences between requested and assigned final grades, where, for ease of aggregation, letter grades have been converted to numeric grade points according to our institution's scale (A = 4.0, A- = 3.7, B+ = 3.3, etc.). Overall, 76% of students received the grade they requested, and 92% received a grade within a half-letter grade (±0.5) of their request.

5.3.1 Grade Inflation and Student Effort. A potential concern for some institutions or instructors may be final grade inflation caused by the ability for students to redo assignments. In practice, it is straightforward to adjust grade target criteria to take this factor into account if desired, essentially baking higher expectations into the defined grade targets.

To evaluate the effect on average grades of our grading scheme, we compared students in the three courses using the new grading scheme with students in these courses the previous time they were taught using a traditional grading scheme. Each course was modified very little from its previous offering, with the exception of switching to the proposed grading scheme. In all cases the courses were taught by the same instructor. In total, there were 100 students combined in traditionally graded sections, and 97 students experiencing the new grading scheme. The GPA increased just 3% from traditional grading to new grading. The change was not found to be significant (p=0.377).

A follow-up question is what effect the new grading scheme has on student effort. To measure the effect, we looked at our university end-of-course surveys, administered to all students for every course at the end of each semester. This survey includes the question: "I have consistently put a lot of effort into meeting course objectives," measured on a six-point scale, where 6 is "strongly agree." We compared responses from students in the three courses in the most recent traditionally graded offering with responses from students in the new grading scheme. 42 students completed the survey in a traditionally graded format in spring 2020, fall 2020, or spring 2021, and 37 students took the survey in an alternatively graded format (fall 2021 or spring 2022). On average, students scores increased 11%. This change was found to be statistically significant (p=0.001). Figure 4 compares the distribution of scores using a kernel density estimate (KDE) from students experiencing each format. It is clear that, although the other aspects of the courses remained very similar from one semester to the next, students felt that they worked harder under the new grading scheme.

5.4 Student Comments

In addition to the quantitative questions, our post-survey included the free-text prompt: "What other thoughts do you have about the course? Were there parts of the course that were especially effective and should be continued? What things could be improved for next time?"

Overall, students were overwhelmingly positive about the new grading scheme, with many comments such as "I really liked the grading style. I felt like it helped me not stress but still held me accountable for work." Several comments related reduced anxiety to the ability to redo an assignment: "... helped me to stress less about one part of something, and focus more on learning the material for a retake." Other students noted that not having numeric grades helped reduce anxiety: "...took away the compulsive pressure I put on myself to get the highest score even though the knowledge matters more than the GPA."

Another theme among student comments was a feeling of greater control over course outcomes: "The system made me feel in control of the amount of effort I wanted to put in." Some student comments spoke to increased intrinsic motivation: "[not having] numerical grades helped motivate students to actually learn the material and not worry so much about a grade." And one student noted that their opinion about the grading scheme changed over the course of the semester: "The grading portion was very new and at first I did not like it. It overwhelmed me not to know my current standing in the

class, but I learned to focus on the assignments instead of just the grade."

6 LIMITATIONS

Our study includes several limitations. We measured changes in student motivation from start to end in each alternatively graded course. This strategy does not compare changes in student attitudes across different grading schemes, and does not account for changes that may be due to factors unrelated to the grading scheme, such as how the classes are taught or how often students came to office hours. We focused on three upper-level elective courses taught by one instructor over one academic year. The extent to which these results would generalize more broadly to other courses, cohorts of students, or instructors still needs to be established in future work. Finally we compared student perceptions of consistent effort under the proposed grading scheme with students in the traditionally graded prior offering of each course. These prior offerings in some cases incorporated hybrid in-person/online formats due to pandemic-related changes that may also have affected the results.

7 CHALLENGES

Putting the new grading scheme into practice, we had several concerns. For the most part, the anticipated difficulties proved less problematic than anticipated.

Increased Grading Effort. In our experience, the bulk of grading effort involves giving good feedback, and this part remains the same in the absence of numeric grading. However, because the resolution is coarser (e.g., does this assignment sufficiently meet requirements?), less time is wasted worrying about fine-grained decisions (e.g., is this error worth a 5-point or 4-point reduction?), which makes assessment faster and less tedious. Further, while allowing students to redo assignments adds instructor work, most students do not resubmit many assignments, and for us, most of the redos tended to be on lab assignments that were primarily checked off for completion or auto-graded multiple-choice quizzes. Even the larger programming assignments are much easier to assess the second time around because the instructor has provided students with feedback from the first assessment which serves as a checklist of items that can be quickly verified during a reassessment. Unfortunately, the new grading scheme does increase bookkeeping to keep up with redo requests and counts, but a bit of programmatic automation can make this effort easier.

Communicating with/convincing students. At the college level, students have invested many years mastering numeric grading systems and may be reluctant to consider alternatives [6]. While a few students did indicate some initial concern, overall students were very positive about the new scheme. We believe three key elements made the transition smooth. First, we motivated the new scheme with pedagogical research anticipating the benefits we expected the students to experience. Second, we laid out clear criteria showing how student achievement would map to final course grades. And third, several times throughout each course, we facilitated discussions with students about how things were going.

Procrastination. We took a poll at the start of each course asking if students would prefer to have a defined period of time when each assignment could be resubmitted (e.g., within two weeks after initial assessment). Perhaps unsurprisingly, students voted in favor of allowing resubmissions through the end of the course. With some trepidation about the potential volume of assessment work during the last week, we nonetheless agreed to the plan. In practice, while there was a clear uptick of resubmissions late in the semester, students were more often responsible about resubmitting in a reasonable time. As a bonus, for the first time in our experience, the number of requests for extra credit late in the semester was zero.

Decreased Student Effort. We worried that, if the goal on an assignment is simply to meet the requirements, with no special recognition given for excelling (i.e., getting *all* the points), perhaps students would lack the motivation to push themselves. In fact, students reported more consistent effort, on average, than in prior semesters (see Section 5.3.1). We also observed that, given the opportunity to do optional "challenge" problems on assignments, more students attempted these extra problems even though they were explicitly not required. We also observed that students seemed more inclined to experiment during in-class work ("what happens if I try this?"), which we attribute to a more relaxed attitude due to their confidence about meeting the requirement of completing the assignment as opposed to trying to maximize a numeric score.

8 CONCLUSIONS AND FUTURE WORK

Over the past year, we implemented a new grading scheme designed to improve student motivation by decreasing the focus on grades. The new system eliminates numeric grades, provides students with opportunities to redo assignments that do not completely meet requirements, and gives students input into their final course letter grade. In three courses using the new grading scheme, students reported a significant increase in intrinsic motivation, self-efficacy, and feeling of control over their own learning. At the same time, students reported applying more consistent effort to their learning compared with students in traditionally graded versions of the courses. Survey comments indicate that students enjoyed the new approach and felt less anxiety than in other courses with traditional grading.

From an instructor standpoint, we find that the proposed changes have made the grading process more enjoyable by removing the necessity of precise numeric score decisions. Further, the number of learning-focused, instructor-student conversations about homework feedback increased dramatically, while the number of conversations about homework grades decreased to almost zero. More students came to office hours with specific questions leading to improved learning outcomes ("I realized after the last quiz that I never really understood precision and recall; can we go over some examples?")

For the future, we plan to apply our ungrading approach to additional courses. We are interested in increasing technology support for auto-generating programming problems and quiz questions to enable second chance attempts to scale to larger numbers of students. We are also interested to investigate the effect of our grading scheme on academic dishonesty. We believe that improved intrinsic motivation and self-efficacy should lead to improved academic integrity.

REFERENCES

[1] Andrew Berns. 2020. Scored out of 10: Experiences with binary grading across the curriculum. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 1152–1157.

[2] Andrew Berns and J Philip East. 2020. Toward a Pedagogy Compatible with Equitable Grading: Tenets, Difficulties, and Suggestions. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 1358–1358.

[3] Benjamin S Bloom. 1968. Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment* 1, 2 (1968), n2.

[4] Susan Debra Blum. 2020. *Ungrading: Why rating students undermines learning (and what to do instead)*. West Virginia University Press.

[5] Susan M Brookhart, Thomas R Guskey, Alex J Bowers, James H McMillan, Jeffrey K Smith, Lisa F Smith, Michael T Stevens, and Megan E Welsh. 2016. A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research* 86, 4 (2016), 803–848.

[6] Sarah Brown and Victoria Chávez. 2022. Communicating Alternative Grading Schemes: How to Shift Students' Attention to Their Learning from Grades. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*. Association for Computing Machinery, New York, NY, USA, 1189–1189.

[7] Ruth Butler. 1988. Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British journal of educational psychology* 58, 1 (1988), 1–14.

[8] Gary Chu. 2020. The Point-less Classroom: A Math Teacher's Ironic Choice in not Calculating Grades. In *Ungrading: Why rating students undermines learning (and what to do instead)*, Susan Debra Blum (Ed.). West Virginia University Press, Chapter 10, 161–170.

[9] A Duryee Crooks. 1933. Marks and marking systems: A digest. *The Journal of Educational Research* 27, 4 (1933), 259–272.

[10] Shannon Duvall, Scott Spurlock, Dugald Ralph Hutchings, and Robert C Duvall. 2021. Improving Content Learning and Student Perceptions in CS1 with Scrumage. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 474–480.

[11] Joe Feldman. 2018. *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin Press.

[12] Dan Garcia, Connor McMahon, Yuan Garcia, Matthew West, and Craig Zilles. 2022. Achieving "A's for All (as Time and Interest Allow)". In *Proceedings of the Ninth ACM Conference on Learning @ Scale*. Association for Computing Machinery, New York, NY, USA, 255–258.

[13] Fran Hassencahl. 1979. Contract grading in the classroom. *Improving College and University Teaching* 27, 1 (1979), 30–33.

[14] Alfie Kohn. 1999. From degrading to de-grading. *High school magazine* 6, 5 (1999), 38–43.

[15] David L. Largent. 2022. Instantiating Specifications Grading in Computer Science Courses. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*. Association for Computing Machinery, New York, NY, USA, 1118.

[16] David L. Largent, Christian Roberson, and Carlo Sgro. 2021. Transform Your Computer Science Course with Specifications Grading. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 1355.

[17] Michelle Meadows and Lucy Billington. 2005. A review of the literature on marking reliability. *London: National Assessment Agency* (2005).

[18] Grace M Mirsky. 2018. Effectiveness of specifications grading in teaching technical writing to computer science students. *Journal of Computing Sciences in Colleges* 34, 1 (2018), 104–110.

[19] Tamera B Murdock and Eric M Anderman. 2006. Motivational perspectives on student cheating: Toward an integrated model of academic dishonesty. *Educational psychologist* 41, 3 (2006), 129–145.

[20] Serena Nicoll, Kerrie Douglas, and Christopher Brinton. 2022. Giving Feedback on Feedback: An Assessment of Grader Feedback Construction on Student Performance. In *LAK22: 12th International Learning Analytics and Knowledge Conference*. 239–249.

[21] Linda B Nilson. 2015. *Specifications grading: Restoring rigor, motivating students, and saving faculty time*. Stylus Publishing, LLC.

[22] Claudia Ott, Brendan McCane, and Nick Meek. 2021. Mastery Learning in CS1-An Invitation to Procrastinate?: Reflecting on Six Years of Mastery Learning. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*. Association for Computing Machinery, New York, NY, USA, 18–24.

[23] Leah Perlmutter, Jayne Everson, Ken Yasuhara, Brett Wortzman, and Kevin Lin. 2022. Reading Between the Lines: Student Experiences of Resubmission in an Introductory CS Course. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*. Association for Computing Machinery, New York, NY, USA, 1137.

[24] Paul R Pintrich et al. 1991. A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). (1991).

[25] Christopher K Riesbeck. 2020. Critique-Driven Learning and Assessment. In *Ungrading: Why rating students undermines learning (and what to do instead)*, Susan Debra Blum (Ed.). West Virginia University Press, Chapter 8, 123–139.

[26] Christian Roberson. 2018. Techniques for using specifications grading in computer science. *Journal of Computing Sciences in Colleges* 33, 6 (2018), 192–193.

[27] Kevin R. Sanft, Brian Drawert, and Adam Whitley. 2021. Modified Specifications Grading in Computer Science: Preliminary Assessment and Experience across Five Undergraduate Courses. *J. Comput. Sci. Coll.* 36, 5 (jan 2021), 34–46.

[28] Jeffrey Schinske and Kimberly Tanner. 2014. Teaching more by grading less (or differently). *CBE—Life Sciences Education* 13, 2 (2014), 159–166.

[29] Ellen Spertus and Zachary Kurmas. 2021. Mastery-based learning in undergraduate computer architecture. In *2021 ACM/IEEE Workshop on Computer Architecture Education (WCAE)*. IEEE, 1–7.

[30] Hugh Taylor. 1980. Contract Grading. (1980).

[31] Ella Tuson and Tim Hickey. 2022. Mastery Learning and Specs Grading in Discrete Math. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*. Association for Computing Machinery, New York, NY, USA, 19–25.

[32] Ursula Wolz, Gail Carmichael, Dan Garcia, Bonnie MacKellar, and Nanette Veilleux. 2020. *To Grade or Not To Grade*. Association for Computing Machinery, New York, NY, USA, 479–480.