Alternative Grading Systems: Quantitative Comparisons

Dan S. Myers
dmyers@rollins.edu
Rollins College
Winter Park, Florida, USA

ABSTRACT

Interest in alternatives to traditional grading has increased in recent years, but instructors exploring new grading approaches have few guides to the trade-offs between different systems. This paper compares a set of alternative grading approaches by regrading real student work from three different computer science classes according to six grading systems. Comparing the resulting letter grade distributions yields insights into how different strategies for assigning credit and determining final letter grades can affect student outcomes. Hybrid systems that combine specifications and labor-based grading are promising, reducing grade inflation while also supporting the success of traditionally underrepresented students. These results have practical value for instructors creating new grading systems and establish promising directions for future research on assessment in computer science.

CCS CONCEPTS

• **Social and professional topics** → **Student assessment**; **Computer science education**.

KEYWORDS

Grading, Specifications grading, Assessment, Education

ACM Reference Format:
Dan S. Myers. 2022. Alternative Grading Systems: Quantitative Comparisons. In *Proceedings of PREPRINT*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

For most instructors, grading is a necessary but unappealing fact of life. Evaluating students is a key element of our educational system, but our grading practices are often guided by tradition and the need for efficiency, rather than careful reflection and research. Recent years, though, have seen increased interest in both the philosophy and mechanics of grading, along with arguments that new grading approaches will enhance student learning and promote equity. [2, 8–10, 13]. In the wake of COVID, these arguments have extended to "ungrading," and minimizing or even eliminating traditional grades [2, 18]. However, instructors interested in new assessment methods have few guides to the trade-offs between different grading approaches. Even when instructors embrace arguments for new systems, concerns about rigor and fairness present barriers to their implementation. This paper addresses those gaps in the literature by considering a practical question: How do alternative grading systems affect student outcomes in computer science classes?

This work empirically evaluates popular alternative grading methods by regrading real student work from three different CS classes under six different grading systems. Observing changes in the resulting letter grade distributions reveals how different ways of awarding credit and assigning letter grades affect outcomes, both at the level of entire classes and for important student demographic subgroups. In addition to illustrating the trade-offs between different grading approaches, this study supports three conclusions:

(1) Reasonable systems that seem student-friendly can actually have a negative impact on low-performing students and increase failure rates compared to traditional grading.
(2) Systems with a labor-based component, where the core coursework establishes a base grade and optional additional work is required to earn the highest grades, reduce the proportion of A grades, but tend to improve grades for motivated B and C-level students.
(3) Hybrid systems, combining specifications and labor-based grading, close the average grade gap between Black/Hispanic and other students in our evaluation group. These systems award fewer A's overall (they are anti-inflationary), but higher percentages of women and Black/Hispanic students earn A's under hybrid grading than under traditional systems or pure specifications grading. Further, hybrid grading systems do not increase failure rates for any student group.

The rest of this paper is organized as follows. Section 2 summarizes the courses and student work in our data set. Section 3 describes six literature-based example grading systems, and Section 4 compares the letter grade distributions obtained under each system for both complete classes (Section 4.1) and student demographic subgroups (Section 4.2). Section 5 discusses limitations, and Section 6 discusses open problems identified by this research that are promising directions for future work.

2 METHODS

This paper investigates the differences among a set of representative grading systems by regrading a set of real student work taken from three CS courses that I taught in 2020-2021, focusing on the differences in the final A-F letter grade distributions produced by each system. The following research questions are rooted in my experience using specifications grading for three years in more than a dozen computer science classes:

- (RQ1) Are there meaningful differences in the grade distributions obtained under different grading systems?

PREPRINT, 2022,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Name	Students	Women	BIPOC
Bridge	23	8 (35%)	11 (48%)
Simulation	19	5 (26%)	6 (32%)
Networks	20	4 (20%)	6 (30%)

Table 1: Course names and demographics

- (RQ2) Does the choice of grading system affect the outcomes experienced by different student subgroups?
- (RQ3) Do more complex hybrid systems yield different results from simpler specifications-based systems?

The raw data consists of all student work submitted in three CS courses from the 2020-2021 academic year:

- A “bridge” course taken after CS1 that focuses on object-oriented programming in Java and software development.
- A junior-level course in probability and simulation.
- A senior-level course in computer networks.

Student demographics are shown in Table 1. The three courses cover different topics and levels, but together are representative of core coursework in the CS curriculum. Section 4 will show that the different grading systems have similar effects on all three courses.

The three courses were originally offered under a hybrid specifications grading system (referred to as FINE HYBRID in Section 3) using three types of assessments:

- Five to seven homework assignments, such as problem sets and programming projects.
- Four to six quizzes spaced throughout the semester — there were no comprehensive midterm or final exams.
- Three optional “challenge projects” that were not part of the core coursework but could be completed to earn the highest grades or make up for unsatisfactory work.

The complete data set contained 17 assignments (356 individual submissions), 16 quizzes (326 submissions), and 9 challenge projects (114 submissions). Each assignment came with a list of specifications, such as automated tests or necessary features, that were required for full credit. Satisfactory completion of 90% or more of the core coursework (assignments and quizzes) earned a base grade of B; failure to meet the B standards reduced this base grade. To earn grades higher than B or make up for missed coursework, students could complete the challenge projects: each one boosted the final course grade by one part of a letter (e.g., B to B+).

Regrading of all student work took place after the conclusion of the 2020-2021 academic year. The regrading process evaluated all student work and assigned final grades in each class according to each of the six systems described in Section 3.2. Each piece of work was regraded separately for each system, as opposed to using the grade under one system as a proxy for the others. Section 5.1 discusses the relationship between the regrading methodology and the conclusions and generalizability of this study in more detail.

3 GRADING SYSTEMS

3.1 Related Work

The literature on alternative grading discusses a number of different mechanisms and philosophies that frequently overlap with each other. The alternative systems examined in this paper draw primarily from the specifications and labor-based grading approaches. Specifications grading systems assess work using only two categories: a submission is either satisfactory and receives full credit, or unsatisfactory and receives no credit [13]. Every assignment comes with a set of formal specifications — for example, a rubric or set of automated tests — that clearly state the requirements that students must meet. “Satisfactory” generally refers to work that would earn a B under a traditional system, so specs grading is primarily concerned with affirming competence, rather than ranking or making fine distinctions between students. Nilson argues that specs grading ties grades to demonstrated learning [13, 20] and can be used to support backwards course design [21]. Labor-based systems emphasize engagement and continuous improvement by assigning grades based on the total amount of work students complete during the course [4, 6, 7, 9]. Students who do more — problems, projects, etc. — earn higher grades than those who do less. Instructors can set required quality guidelines for each assignment, but acceptable completion, rather than points, is the basis for grades.

Several computer scientists have submitted experience reports describing their alternative grading systems [11, 12, 15, 16, 19]. In a recent detailed work, Berns discussed “binary” grading in three courses [1]. His system scores entire assignments as 0/1, allows unlimited retakes, and awards final grades based on the number of satisfactory assignments, so it combines elements of both specs and labor-based grading. Riesbeck describes a similar “critique-driven” approach in his AI Programming course that he has used since 1997: students submit solutions to problems, receive feedback, and resubmit until their solutions are satisfactory [14]. Both Berns’ and Riesbeck’s systems fit into the framework of mastery grading, where students are allowed to resubmit work until they achieve “mastery” of the content [3]. Sanft et al. describe a modified specs system that allows resubmissions but reduces the maximum assignment score for each attempt. They evaluate the grading times and student perceptions of this system across five CS classes; results show time savings of approximately 50% on most specs graded assignments compared to traditional points-based grading [17].

3.2 Six Grading Systems

This section describes our six different grading approaches: one points-based system, two binary specifications systems, a labor-based system, and two hybrid systems. Although it’s unlikely that another instructor would adopt any of these approaches as-is, together they represent a set of strategies that faculty might draw from when designing a new grading system. The systems are distinguished by how they evaluate coursework and how those evaluations map to letter grades. The points-based and binary specs systems evaluate only the core coursework (assignments and quizzes) and do not use the optional challenge projects. The other three systems include a labor-based element, where the extra challenge projects are used as optional work beyond the required core.

Points-based system. A straightforward points-based system is the baseline for all of our comparisons. The POINTS system scores all core coursework (assignments and quizzes) out of 100, awarding 0 for missed work. Final grades are calculated by an equally-weighted

Name	Description	Scoring	Letter Grades	Challenge Projects
POINTS	Points-based with numerical scale	Percentage out of 100	Numerical average of assignments and quizzes	Unused
COARSE BINARY	Binary evaluation of entire assessments	0/1 credit for assignments and quizzes; high-C to B threshold	Fraction of satisfactory assessments	Unused
FINE BINARY	Binary evaluation of individual questions or project features	0/1 credit for each question or feature	Fraction of total binary points	Unused
LABOR	Grades are based on total completed work	"Good effort" showing application of the course material and overall correctness	$N-1$ assessments earns a base grade of B, $N-2$ earns a C, etc.	Each project raises final grade by one part of a letter (e.g., B to B+)
COARSE HYBRID	Hybrid combining COARSE BINARY and LABOR	0/1 credit for assignments and quizzes	90% of binary points earns a base grade of B	Each project raises final grade by one part of a letter (e.g., B to B+)
FINE HYBRID	Hybrid combining FINE BINARY and LABOR	0/1 credit for each question or feature	90% of binary points earns a base grade of B	Each project raises final grade by one part of a letter (e.g., B to B+)

Table 2: Summary of the six grading systems

average of all assignment and quiz grades with an average of 90% required for an A, 80% for B, and so forth. This system does not include "extra credit," so it does not incorporate the challenge projects.

Specifications systems. Section 4 evaluates two binary versions of specs grading, similar to [1, 14]. The first, COARSE BINARY, assigns 0/1 credit to entire assignments or quizzes, with no possibility of partial credit. The second, FINE BINARY, is a finer-grained system, where each individual problem, quiz question, or project feature is evaluated independently and awarded 0/1 credit. This system allows students to earn credit for some parts of an assignment even if others are incomplete. In both systems, letter grades are determined by the total fraction of 0/1 points, with 90% required for an A, 80% for a B, etc.

Labor-based system. Our LABOR system is based on one developed by Asao Inoue for his writing classes [9]. Completion of $N-1$ pieces of the core coursework with good effort earns a base grade of B, $N-2$ earns a C, and so forth. The challenge projects function as the extra labor: each satisfactory challenge project raises the final course grade by one part of a letter (e.g., B to B+). Inoue connects his approach to antiracist teaching and the challenge of designing a fair grading system when students have unequal prior experience with the subject — a familiar problem for CS1 instructors.

Hybrid systems. Our final two systems combine a specs-graded core body of work with a labor-based component required for the highest grades. The first, FINE HYBRID, is the original grading system I used during the 2020-2021 academic year, described in Section 2. This system grades student work as in the FINE BINARY specs system, but earning 90% of the total binary points over the semester earns a base grade of B. The challenge projects are used to award grades above B or make up for unsatisfactory work, as in the LABOR system. The second hybrid system, COARSE HYBRID, uses the same approach, but awards 0/1 credit to entire assignments, as in the COARSE BINARY specs system.

4 RESULTS

This section presents comparisons of the grade distributions obtained under the six different systems. The comparisons focus on how the A-F letter grade distributions differ between the traditional POINTS system and each of the alternative approaches. Although all of the systems are capable of awarding fractional plus/minus grades, these results award only the base A-F letters, with no fractional adjustments. Table 3 compares the grade distributions and mean class grade points for the Bridge, Simulation, and Networks courses across all six grading systems. The GPA scores have been calculated by converting each student's A-F letter grade to a numerical value using a standard 4.0 scale, then averaging the grade point scores for each class.

4.1 Comparisons

4.1.1 Binary specifications grading. Let's first consider the two pure specifications systems: COARSE BINARY and FINE BINARY. The COARSE BINARY system, which assigns 0/1 credit to entire assessments, yields the highest grades of any system. FINE BINARY, which assigned 0/1 credit to each individual question or project feature, had the opposite effect, yielding the lowest average GPA and highest failure rates.

The first result is intuitive: the COARSE BINARY system awards full credit to any assignment that passes the satisfactory threshold, so grades above the high-C to B boundary are pulled up to A's. Figure 1 shows the changes in individual student grades under the two systems. Each row represents grades students earned under the POINTS system; columns represent grades under the COARSE BINARY system. Values on the diagonal report the numbers of students who earned the same grades under both systems and off-diagonal values

Course	System	GPA	A	B	C	D	F
Bridge	POINTS	2.78	.48	.13	.17	.13	.09
	COARSE BINARY	3.13	.69	.04	.04	.13	.09
	FINE BINARY	2.65	.43	.17	.13	.13	.13
	LABOR	2.73	.43	.26	.04	.13	.13
	COARSE HYBRID	3.03	.52	.21	.13	.04	.08
	FINE HYBRID	2.78	.39	.26	.17	.08	.08
Simulation	POINTS	3.10	.53	.32	.00	.05	.10
	COARSE BINARY	3.36	.79	.05	.00	.05	.10
	FINE BINARY	2.84	.47	.21	.16	.00	.15
	LABOR	3.00	.52	.26	.05	.00	.15
	COARSE HYBRID	3.05	.52	.26	.05	.05	.10
	FINE HYBRID	2.89	.47	.21	.16	.05	.10
Networks	POINTS	2.95	.60	.05	.15	.10	.10
	COARSE BINARY	3.00	.50	.25	.10	.05	.10
	FINE BINARY	2.70	.60	.05	.05	.05	.25
	LABOR	2.80	.50	.25	.00	.05	.20
	COARSE HYBRID	3.00	.50	.25	.10	.05	.10
	FINE HYBRID	2.85	.50	.20	.05	.15	.10

Table 3: Distributions of letter grades and GPAs obtained under each of the six grading systems

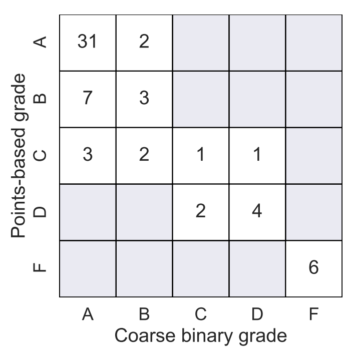

Figure 1: Grade changes: POINTS vs. COARSE BINARY

report students who received different grades under COARSE BINARY than they did under POINTS. The figure aggregates the results from all three courses—trends at the individual course level were similar. Seven total students received B's under POINTS but moved up to A's under COARSE BINARY, and three students moved from C's all the way up to A's. Only three total students earned a lower grade under COARSE BINARY than under POINTS.

Figure 2 shows that individual grade changes are consistently worse under the FINE BINARY system: only one student moved to a higher grade and two-thirds of D students moved down to failing. These changes are due to a tendency to award at least some partial credit to incorrect answers under the POINTS system, which is not possible under FINE BINARY. The loss of question-level partial credit is particularly harmful to students in the low-C and D range. Overall, 35% of students saw their grades drop by at least half a letter grade under FINE BINARY compared to POINTS.

Insights. Pure 0/1 grading is easy to implement and efficient. Binary grading of entire assignments is a good for for mastery-based

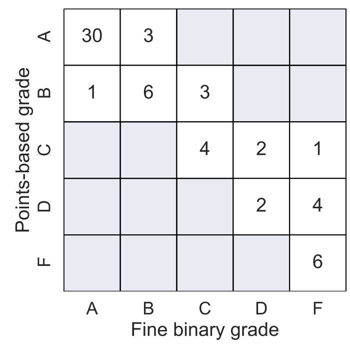

Figure 2: Grade changes: POINTS vs. FINE BINARY

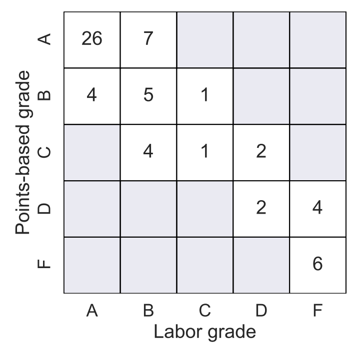

Figure 3: Grade changes: POINTS vs. LABOR

systems, where standards for credit are high but students are given feedback and multiple attempts to meet those standards [3]. Finer-grained specs systems make it easy to combine different types of assignments and exams into one system, but can hurt low-performing students if they lose question-level partial credit that they could have earned under a points-based system.

4.1.2 Labor-based grading. The LABOR system awards 0/1 credit to entire assessments, like COARSE BINARY but with even more relaxed standards, so we might expect it to lead to even higher grades. In fact, results show that labor-based grading actually led to a small *decrease* in both GPA and the percentage of A's in all three classes (Table 3). In practice, earning credit under the "good effort" standard was closely correlated to earning 0/1 credit under COARSE BINARY, which makes it more surprising that LABOR did not lead to a similar boost in grades. The individual grade changes (Figure 3) show that many students who earned A's under POINTS based on only the core coursework dropped down to B's because they did not complete the challenge projects required for an A, but some B and C students were able to move up to higher grades because they did complete multiple challenge projects. Note, however, that low-performing students struggled. The LABOR system reduces letter grades aggressively after one missed assignment, more so than the binary specs systems. Students on the D/F boundary generally missed multiple assignments and did not complete the extra challenge projects, so they did worse under LABOR than they would have under the POINTS

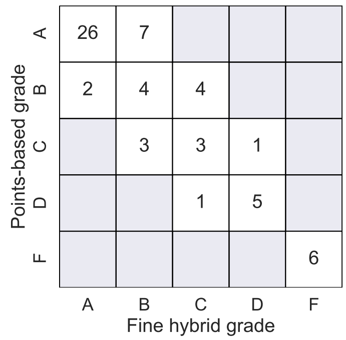

Figure 4: Grade changes: POINTS VS. FINE HYBRID

system, which allows for more partial credit and a more gradual decrease in grades.

Insights. Labor systems have an anti-inflationary effect on A grades, while still allowing B and C students to raise their grades if they put in extra work. Labor systems seem student-friendly, but they can punish students who fall behind and can't complete extra work to recover their grades.

4.1.3 Hybrid systems. The two hybrid systems combine a specs-graded core with extra work required to earn the highest grades. Table 3 shows that the COARSE HYBRID system generally raised grades and the FINE version generally reduced them, in line with their binary counterparts, but that GPA and grade distributions remained closer to the POINTS system. Figure 4 plots the individual grade changes for the FINE HYBRID system. The system shows an anti-inflationary effect on A and B grades, but only one student at the C or D level moved down to a lower grade. This is encouraging: the specs-based component lowers the base grade more gradually than LABOR, so low-performing students remain at the C or D level and don't drop down to failing. The COARSE HYBRID system (not shown) has similar results: its grades are generally higher than POINTS, but the system does not boost grades as much as COARSE BINARY.

Insights. These hybrid systems demonstrate an anti-inflationary effect at the high grades without harming low-performing students. The general concept of a core body of work that establishes a B or C grade supported by optional assignments is a promising model, but more complex and harder to administer than our other systems.

4.2 Student Subgroups

Figure 5 plots the average grade points for women ($N = 17$) and men ($N = 45$) under the six grading systems, combining data from all three courses (as before, the individual courses exhibit similar behaviors). The plot shows that women outperform men under every system by approximately .70 grade points: a significant gap, but in line with the overall academic profile of the institution. This gap is consistent across all six systems.

Figure 6 compares GPAs for Black/Hispanic-identifying students ($N = 19$) and other students, which includes those identifying as White, Asian, and Middle Eastern/North African ($N = 43$).

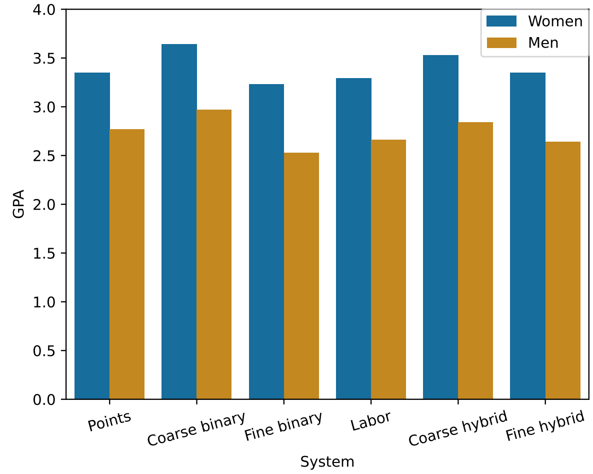

Figure 5: GPA comparisons: women and men

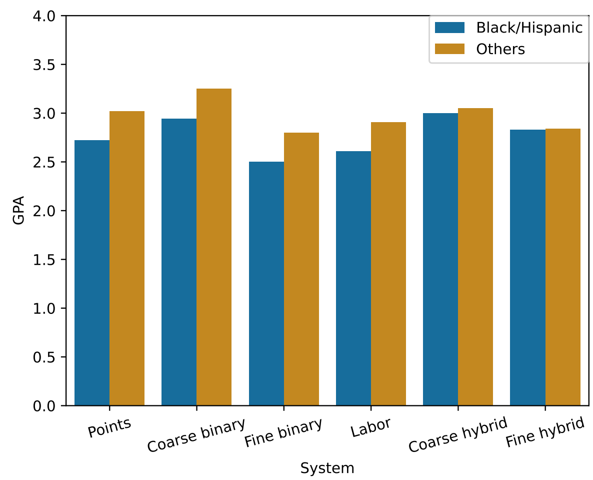

Figure 6: GPA comparisons: Black/Hispanic and others

The decision to aggregate students into these groupings is motivated by the relatively small size of the Black ($N = 4$) and Asian ($N = 6$) groups. Four grading systems show a consistent gap of about .30 grade points, but the two hybrid systems achieve almost equal GPAs between the two subpopulations. Compared to POINTS, Black/Hispanic students gained an average of .11 grade points under FINE HYBRID (2.72 to 2.83) with a loss of .18 for the "Others" category (3.02 to 2.84).

Table 4 shows the percentages of students in each demographic subgroup that attained A and F grades under each system. The results support the conclusions from the previous section regarding the relative dynamics of the different systems.

- LABOR does increase the percentage of A's for both women and Black/Hispanic students, but disproportionately increases failure rates for both groups. Notably, 22% of Black/Hispanic

System	Women %A	Women %F	Men %A	Men %F	Black/Hispanic %A	Black/Hispanic %F	Non-Black/Hispanic %A	Non-Black/Hispanic %F
POINTS	65	0	49	13	44	5	56	11
COARSE BINARY	82	0	60	13	61	5	68	11
FINE BINARY	64	11	44	20	44	22	52	16
LABOR	70	11	40	17	55	22	45	13
COARSE HYBRID	70	0	44	13	55	5	50	11
FINE HYBRID	65	0	37	13	50	5	43	11

Table 4: Percentage of each subgroup attaining A and F grades under each system

students received failing grades under LABOR, compared to only 5% under the POINTS and HYBRID systems.

- Under FINE HYBRID, the percentage of failing grades did not increase for any subgroup, but the percentage of A's dropped for both men and non-Black/Hispanic students. Black/Hispanic students saw a small increase in their percentage of A's under FINE HYBRID.
- COARSE HYBRID led to higher grades than POINTS on average (per Table 3), but still had an anti-inflationary effect on men and non-Black/Hispanic students.

In fact, the percentage of White men earning A's under both hybrid systems was lower than under FINE BINARY, which had the lowest overall GPA among the six systems. White men in our group were less likely to undertake the additional challenge projects to raise their grades than were women and Black/Hispanic students.

Insights. Among these students, hybrid systems close the average grade gap between Black/Hispanic and other students. In general, women and Black/Hispanic students we observed were more willing to undertake the additional challenge work, which boost the percentage of A's earned by the highest-performing students in those groups compared to the POINTS system. Both hybrid systems have an anti-inflationary effect on the grades of White men, but do not increase the failure rate for any group. We cannot conclude, at this time, whether this effect will generalize beyond this set of students, but it does represent a promising direction for future research on equitable grading.

5 DISCUSSION AND LIMITATIONS

The results of this study are best interpreted as examples of how changes in grading mechanics may affect student outcomes within the context of an individual instructor's grading practice. Our regrading results are subject to two key limitations. First, the scores and letter grades obtained under each grading system were based on my own evaluations and individual judgment. Other instructors might, of course, make different decisions about how to award credit or score assignments, which could lead to different grade distributions. Second, although I did anonymize work to the extent allowed by our course learning management system, scoring the same piece of work multiple times means that the regrading process could never be completely blind.

We would expect these two facts to increase correlation among the grading systems — that is, my individual judgment and the repetitive regrading process might have made grades more similar than they would have been in a blinded, independent regrading scheme. Even with those limitations, however, we still observe meaningful differences between the systems. Correlation would be a larger concern if we observed *no* differences. It is possible, though, that other differences were missed that would have emerged under a fully blind grading process.

Further, our sample sizes and student populations are not large or representative enough to support generalized conclusions. In particular, the observed effect of hybrid grading systems closing the performance gap between Black/Hispanic and other students is encouraging, but needs to be supported by a large-scale controlled trial before we can conclude if it extends to CS students in general.

Finally, although our six systems are representative of popular grading strategies discussed in the literature, there are some approaches that couldn't be tested in this format. In particular, because the original courses used fixed deadlines, we could not evaluate the impact of revision or mastery-based grading on outcomes. Likewise, we could not consider contract-based approaches that map letter grades to distinct bundles of assignments[5, 13].

6 CONCLUSION

This paper demonstrates that grading system mechanics can have a meaningful impact on student outcomes beyond simply raising or lowering a course's average grade level. Therefore, we should consider grading and assessment practices an important element of pedagogy, and one that has the potential to influence student success and equity in our classes. Our results raise a number of interesting possibilities for future research. The impact of hybrid grading systems on student achievement gaps is promising, but needs to be validated by a larger-scale study. Further, a larger study would also allow us to examine the challenges of deploying alternative grading methods at scale and the related issues of intergrader agreement and reliability. There is also the possibility of examining the intersection of different grading strategies with autograding, which can build early formative feedback and opportunities for revision into every assignment. Finally, the questions considered in this study are important to other disciplines, not only computer science. Therefore, there are abundant opportunities for interdisciplinary collaborations to improve grading across the STEM curriculum.

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*. 1152–1157.

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

[3] Clark, David [n.d.]. *Mastery/Pointless Grading FAQ.* https://docs.google.com/document/d/1oWBOxRhU3kqizhJcbSYFc-33p_HyftA4FYh4zI6-

ZUA/,lastaccessed=

[4] Jane Danielewicz and Peter Elbow. 2009. A unilateral grading contract to improve learning and teaching. *College Composition and Communication* (2009), 244–268.

[5] Davidson, Cathy 2011. *Contract Grading + Peer Review: Here's How It Works*. Retrieved July 26, 2021 from https://www.hastac.org/blogs/cathy-davidson/2011/01/03/contract-grading-peer-review-heres-how-it-works

[6] Peter Elbow. 1993. Ranking, evaluating, and liking: Sorting out three forms of judgment. *College English* 55, 2 (1993), 187–206.

[7] Peter Elbow. 1997. Grading student writing: Making it simpler, fairer, clearer. *New directions for teaching and learning* 1997, 69 (1997), 127–140.

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

[9] Asao B Inoue. 2019. *Labor-based grading contracts: Building equity and inclusion in the compassionate writing classroom*. WAC Clearinghouse.

[10] Alfie Kohn. 2011. The case against grades. *Educational Leadership* 69, 3 (2011), 28–33.

[11] James W McGuffee, David L Largent, and Christian Roberson. 2019. Transform your computer science course with specifications grading. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 1234–1234.

[12] 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.

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

[14] Christopher Riesbeck. 2020. Critique-Driven Learning and Assessment. In *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)*, Susan D Blum (Ed.). West Virginia University Press, Chapter 8, 123–139.

[15] Christian Roberson. 2017. Applications of Specifications Grading in Computer Science Courses. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 716–716.

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

[17] Kevin R Sanft, Brian Drawert, and Adam Whitley. 2021. Modified specifications grading in computer science: preliminary assessment and experience across five undergraduate courses. *Journal of Computing Sciences in Colleges* 36, 5 (2021), 34–46.

[18] Stommel, Jesse 2018. *How to Ungrade*. Retrieved July 27, 2021 from https://www.jessestommel.com/how-to-ungrade/

[19] Torstein JF Strømme. 2019. Pass/fail grading and educational practices in computer science. In *Norsk IKT-konferanse for forskning og utdanning*.

[20] Barbara E Walvoord and Virginia Johnson Anderson. 2011. *Effective grading: A tool for learning and assessment in college*. John Wiley & Sons.

[21] Grant P Wiggins and Jay McTighe. 2005. *Understanding by design*. ASCD.