# Experience Report: Standards-Based Grading at Scale in Algorithms

Lijun Chen*
lijun.chen@colorado.edu
Department of Computer Science
University of Colorado Boulder

Joshua A. Grochow
joshua.grochow@colorado.edu
Departments of Computer Science & Mathematics
University of Colorado Boulder

Ryan Layer
ryan.layer@colorado.edu
BioFrontiers Institute & Department of Computer Science
University of Colorado Boulder

Michael Levet
michael.levet@colorado.edu
Department of Computer Science
University of Colorado Boulder

## ABSTRACT

We report our experiences implementing standards-based grading at scale in an Algorithms course, which serves as the terminal required CS Theory course in our department's undergraduate curriculum. The course had 200-400 students, taught by two instructors, eight graduate teaching assistants, and supported by two additional graders and several undergraduate course assistants. We highlight the role of standards-based grading (SBG) in supporting our students during the COVID-19 pandemic. We conclude by detailing the successes and adjustments we would make to the course structure.

## CCS CONCEPTS

• **Social and professional topics → Student assessment**; **Model curricula**.

## KEYWORDS

Mastery-Based Grading, Standards-Based Grading, Algorithms

**ACM Reference Format:**
Lijun Chen, Joshua A. Grochow, Ryan Layer, and Michael Levet. 2022. Experience Report: Standards-Based Grading at Scale in Algorithms. In *Proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education Vol. 1 (ITiCSE 2022), July 8–13, 2022, Dublin, Ireland.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3502718.3524750

## 1 INTRODUCTION

Our department's undergraduate Algorithms course serves to meet the in-major Theory requirement. Such courses are standard in undergraduate computer science programs, though they vary substantially in content ranging from implementation-heavy courses in advanced data structures to a more theoretical and proofs-based introduction to algorithm analysis. Our course emphasizes the theoretical content, while incorporating occasional small programming assignments.

This course presents significant and unusual pedagogical challenges, in that it aims to teach theoretical content (the algorithm design techniques and key examples), rigorous problem solving skills (applying and generalizing the techniques), and mathematical communication (i.e., mathematical proof and formalisms). The enrollment is also quite large, ranging from 300-400 students each semester, which limits the amount of individualized attention and detailed feedback that we can provide. The course prerequisites include some Discrete Math course, which can be fulfilled using comparable courses in the Computer Science, Math, or Electrical Engineering departments. Despite this prerequisite, Algorithms is often the first time students are expected to parse or formulate mathematical proof beyond simple numerical induction (e.g., Fibonacci, sum of the first $n$ integers). The culmination of these pedagogical challenges results in a rather steep learning curve for students.

We contrast these challenges with Introduction to Proofs courses in Math departments, which focus on mathematical communication and adopt simpler content (e.g., logic, sets, functions, relations) as the vehicle upon which to practice formulating mathematical proofs. We also contrast our course with upper-division courses in the Math department like Abstract Algebra and Real Analysis, which require prerequisites emphasizing both content and proof-writing skills, and so focus primarily on content. In-major Math courses also tend to be much smaller than Algorithms, usually with fewer than 30 students.

Previous iterations of this course utilized more traditional points-based grading schemes, which failed to effectively communicate to students their strengths and specific areas for improvement. In addition, points-based grading factored each attempt on a given skill into student grades, which effectively punished students for erroneous initial attempts, even if the student went on to demonstrate proficiency later in the course. In our experiences, the drawbacks of points-based grading coupled with the steep learning curve created a course structure that exacerbated large-scale anxiety amongst the students. On the instructional side, implementing points-based grading was time consuming in determining how to consistently award partial credit.

In the Spring 2020 semester, we redesigned the course to incorporate *standards-based grading*, which is the implementation of the

*All authors contributed equally.



*ITiCSE 2022, July 8–13, 2022, Dublin, Ireland.*

ACM ISBN 978-1-4503-9201-3/22/07.
https://doi.org/10.1145/3502718.3524750

principle that *grades should communicate to students their current command of the content and progress with respect to well-defined standards* [25]. In addition to clearly communicating to students their command of course content, we also sought to structure the course grading scheme to reward students for their demonstrated learning prior to the end of the semester, rather than punishing students for failing to (fully) grasp a given skill earlier in the course. We refined our standards-based implementation in subsequent semesters, including Fall 2020, Summer 2021, and Fall 2021.

In this paper, we document our attempts to design an Algorithms course using SBG, including successes and challenges of implementing SBG at scale in a university environment. We also discuss both the challenges we faced in migrating online, as well as the flexibility that SBG enabled us to provide to students. This flexibility was especially impactful during the pandemic, both because students are dealing with more physical and mental health issues, and because of the speed with which university policy changes regarding in-person vs. online vs. hybrid instruction.

### 1.1 Standards-Based Grading

Standards-based grading is a scheme in which students are provided a clearly defined list of learning objectives and multiple attempts to demonstrate proficiency. While ideally students persist until they have demonstrated proficiency for each topic, we note that this is not a requirement for SBG [25]. Final course grades are based upon the number of standards for which students have demonstrated proficiency and the degree of demonstrated learning for each standard [1, 4, 5, 12, 25].

Standards based grading is a popular topic in K-12 education [13, 18, 30], dating back to K-12 educational reform in the 1990s [13]. It has begun to take root in the university setting, most notably in lower-division math [5, 12, 15, 25] and science [1, 16] courses. However, SBG has also been adopted in upper division math courses such as Real Analysis and Abstract Algebra [21, 22], as well as other disciplines such as educational leadership [2, 29] and music [4].

We are only aware of one other instance where SBG has been adopted in Algorithms [24]. Here, the authors organized the content into five core topics: dynamic programming, greedy algorithms, divide-and-conquer, network flows, and computational complexity. Students can earn credit for these topics using periodic assessments throughout the semester, as well as the final exam. In particular, if students demonstrate understanding on a topic early the semester, then they need not attempt that topic on the final exam. The authors observed that this approach resulted in significant improvement in student performance for dynamic programming and proof-writing, which are two of the more challenging topics in the course. Additionally, over two-thirds of students increased their grade by at least 5% by taking the final exam.

Faculty that have implemented SBG report a number of key benefits, both for students and instructors. In particular, students have reported that SBG helped them to focus on understanding the content rather than accruing points [2, 5]. Math instructors have also observed reduced student stress in courses utilizing SBG [5, 12, 23]. On the other hand, student anxiety and discomfort may be initially exacerbated by SBG due to familiarity with more traditional points-based grading systems [16]. Helping students understand the reassessment process is key to quell this anxiety and increase student buy-in [5].

## 2 COURSE CONTENT

We begin by detailing the course content, including details about how the material was presented and the order in which the topics were covered. In subsequent iterations of the course (Summer 2021 onwards), we adjusted both the manner and order in which we covered certain topics in order to help students gain traction, provide better feedback, and reduce the grading workload.

The key topics were as follows. (For a current treatment, see [11].)

- **Proof by Induction & Loop Invariant Proofs**.
- **Asymptotics**. Comparing functions (Calculus techniques), Unrolling and Tree Methods for finding tight asymptotic bounds on recurrences.
- **Analyzing Code**. The goal was to write down the runtime complexity function of a given algorithm, such as iterative algorithms where the loop variables were independent or dependent. For recursive algorithms, students were asked to write down a recurrence for the runtime complexity.
- **Path-Finding**. Breadth-first search, Depth-first search, Dijkstra's Algorithm.
- **Spanning Trees**. Safe & Useless Edges, Prim's Algorithm, Kruskal's Algorithm.
- **Ford–Fulkerson**. Here we emphasized the residual graph, but eventually switched to using the flow network (see Sec. 4).
- **Reductions to Max Flow.** Students were asked to construct or reason about reductions to the Max Flow problem (e.g., Bipartite Matching).
- **Greedy Algorithm Principles**. Examples where greedy algorithms fail to yield optimal solutions (often related to the Interval Scheduling and Making Change problems), Exchange Arguments.
- **Divide & Conquer**. Examples where a divide & conquer algorithm failed to return the correct result, Quicksort, Mergesort
- **Dynamic Programming**.
- **P vs. NP Problem**. Showing problems belong to P or NP; Structure and Consequences of the P vs. NP problem.

Prior to Summer 2021, we covered the course content in the order above, with the first half of the course notably covering proof by induction & loop invariant proofs, and asymptotics. Despite the fact that the key skills (proof by induction, exponentials & logarithms, calculus techniques) are prerequisite content, students report that they find this material to be quite technical. In particular, due to poor early performance on these topics, many students report that they feel set up to fail in the course.

Our treatment of the Ford–Fulkerson algorithm using residual graphs was a pressure point for both students and grading staff. Namely, constructing and updating the residual graph at each iteration of the Ford–Fulkerson algorithm was quite involved. As a result, students frequently made minor careless errors (e.g., omitting an edge) which had significant impact on the final answer.

Additionally, grading student work (even when requiring an initial flow-augmenting path) was quite tedious, with our graders reporting spending 3-5 minutes per submission.

## 3 FIRST ATTEMPT: STANDARDS-BASED GRADING (SPRING AND FALL 2020)

In the Spring and Fall 2020 semesters, we implemented SBG, based largely on the logistics similar to those of Heubrach & Krinsky [9]. We began by identifying 25 content standards, outlined as follows.

- **Proof by Induction & Loop Invariant Proofs (1 standard)**.
- **Asymptotics (3 standards).** Calculus Techniques, Unrolling, Tree Method.
- **Analyzing Code (3 standards)**. Nested independent loops, Nested dependent loops, Recursive algorithms.
- **Path-Finding (2 standards)**. Breadth-first search/Depth-first search, Dijkstra's Algorithm.
- **Spanning Trees (2 standards)**. Safe & Useless Edges, Prim's & Kruskal's Algorithms.
- **Network Flows (2 standard)**. Ford–Fulkerson, Reductions to Max Flow.
- **Greedy Algorithm Principles (2 standards).** Examples where greedy algorithms fail to yield optimal solutions, Exchange Arguments.
- **Hashing (1 standard).**
- **Divide & Conquer (2 standards)**. Examples where a divide & conquer algorithm failed to return the correct result, Quicksort analysis.
- **Dynamic Programming (5 standards).** Identifying subproblems, Writing down recurrence, Designing DP algorithms with one-dimensional lookup table, Designing DP algorithms with two-dimensional lookup tables, Order of subproblems & backtracking.
- **P vs. NP Problem (1 standard).**
- **Participation standard.**

Each content standard had four attempts: a homework, a recitation quiz, a midterm, and the final exam. In order to earn credit for a standard, students had to demonstrate proficiency twice for that standard. Students also had three reassessment attempts via a timed quiz format, provided they first wrote corrections and reflections for their previous attempts (or a correct attempt and explained the material, if the student did not attempt the given standard). Effectively, students had to demonstrate proficiency at least once on a timed assessment. A number of the standards focused on mechanical problems, such as solving a recurrence using the unrolling technique or applying Dijkstra's algorithm to a given graph. Other standards emphasized tasks that required students to generalize concepts or apply techniques in new ways, such as applying a max-flow algorithm to solve new problems, and grappling with the consequences of NP-completeness.

We also included a Participation standard, which was designed to incentivize students to regularly engage with the content rather than relying solely on the midterms and final exam. Our initial intent was to use the Participation standard to incentivize students to regularly attend weekly recitations and work through online multiple choice mini-quizzes, which were different than the recitation quizzes. When we migrated the course to an online format in light of COVID-19, we significantly relaxed the threshold for the Participation standard[1].

Final grades correlated to the number of standards for which students demonstrated proficiency twice, with each standard receiving equal weight. Students did not receive partial credit for a standard if they demonstrated proficiency only once. In order to pass the course (with a C-), students needed to demonstrate proficiency twice on 14 standards. Each additional standard they accrued corresponded to an additional one-third of a letter grade. The cutoff for an A was 21 standards.

Individual problems were graded using five categories: No Attempt, Significant Errors, Progress, Proficiency, Near Perfect. We colloquially referred to these numerically as 0–4, but repeatedly emphasized to students their categorical meaning. For brevity we refer to them here numerically. Students who earned below a 3 on a problem did not receive any credit on a problem. Scores below a 3 were used to provide feedback to students as to their progress on the problem, in addition to written feedback. Students did not receive any additional credit for earning a 4 rather than a 3. Rather, we designed the 3 vs. 4 distinction to recognize exceptional student solutions, as well as to help alleviate student anxiety by signaling a difference between proficiency and near-perfection.

Quizzes and exams each only had one problem per standard. So a student's proficiency score on that question factored directly into their overall course grade. An individual homework assignment, however, often included multiple questions covering a given standard. These scores were then aggregated to determine the overall proficiency score. Students who received at least a 3 on each question for that standard received a 3 or 4 for their overall proficiency score, while students who received below a 3 on every question for that standard received at most a 2 for their overall proficiency score. Similarly, students who received a grade of 0 or 1 on any question for a given standard received at most a 2 for their overall proficiency score. In borderline cases, where students had a mix of 2's, 3's and 4's, we examined their work more closely. Students who demonstrated proficiency on the more challenging questions and had a preponderance of 3's and 4's received an overall proficiency score of at least 3. On the other hand, students who struggled on the more challenging questions or earned primarily scores of 2 received at most a 2 for their overall proficiency score.

**What Worked Well.** Much of what worked well in Spring and Fall 2020 continued to work well in subsequent iterations of the course. We defer this discussion to Sec. 5.

**Pandemic Teaching.** The transition to online teaching in the middle of the Spring 2020 semester, associated with the COVID-19 pandemic, necessitated structural changes in the course to accommodate the online format, as well as to provide flexibility and support to students. Prior to the pandemic, students had weekly homework, a 25-minute closed-book quiz in recitation each week, and a weekly online mini-quiz that contributed to the Participation standard. Homework was to be submitted directly to Canvas, and so little changed in that regard. Each weekly recitation quiz covered

[1]Essentially, the only students who did *not* earn credit for Participation are those who didn't engage with the course at all, even before the shift to remote instruction.

anywhere between one to three standards, with one question per standard. After moving online, each weekly recitation quiz was broken up so that each question was its own timed Canvas assignment, which students could take at any time over a four day period. This allowed students more flexibility to complete the quizzes at their convenience. We calibrated the timing to allow 15 minutes to take the quiz, scaled for students with extra time accommodations, and an extra 15 minutes for students to upload their work to Canvas. In practice, students could use some of the time allocated for uploading work, in order to continue answering the question. Students reported that the online quiz format decreased the pressures associated with a 25-minute in-person quiz. Starting in Fall 2020, students were given 45 minutes per quiz, with an intended 30 minutes to take the quiz and 15 minutes to upload their work to Canvas. We continued this practice even in Fall '21, when the class had both in-person and online sections, because of the ease of logistics.

Students were permitted to use their notes and book, but collaborating with other students and using online tutoring services such as Chegg were prohibited. We also continued requiring students to show their work and document their reasoning. Requiring work and timing the quizzes mitigated the usefulness of posting to message boards or tutoring services, and also allowed us to better ascertain student understanding.

**Participation Standard.** After moving online in Spring 2020, the criterion for earning the Participation standard were opaque. This inspired significant student anxiety and required a significant amount of time on the parts of the instructional staff to manage. During the Fall 2020 semester, the criterion for the Participation standard centered around recitation attendance. As recitations were remote, it incentivized students to log in for attendance credit without engaging in group work activities. Eventually, because of the ongoing pandemic, we removed Participation standards and adjusted the grade scale accordingly.

**Student Buy-In.** A key challenge we faced was in getting student buy-in to the grading scheme. Algorithms is a course in which many students struggle to gain traction. This was exacerbated by the fact that the course has traditionally been front-loaded with respect to the material students find most challenging (i.e., proof by induction, asymptotics). Students also expressed that the Calculus standard covered too much material. Combined with the technical nature of this material, students found attempting this standard to be particularly frustrating. We note that regardless of whether the course employed traditional points-based grading or SBG, student performance early in the semester reflected these difficulties.

With SBG, our goal was to incentivize students to work out the problems completely and correctly. While minor mistakes would make the difference between a 3 and a 4 (both of which counted for full credit), more serious mistakes could result in grades of 1 or 2 for individual problems. While the thresholds for earning a 1 or 2 were quite different, neither score counted towards proficiency. For this reason, students perceived that there was no partial credit under this grading scheme. This observation, coupled with early poor performance, led students to quickly believe they were doomed to failure.

For problem sets, there were often multiple problems for a given standard. Precisely, if there were $d$ problems for a given standard, we defined a mapping $f : \{0, 1, 2, 3, 4\}^d \to \{0, 1, 2, 3, 4\}$ to compute the overall standard score on that assignment. Only the overall standard score was what impacted students' final grades. This mapping was often not a numerical formula; and in particular, rarely the average. Additionally, we did not share any information about our mapping with students, other than that we did not in general use the average to compute overall standard scores. As a result, students struggled to reconcile their scores on each individual problem with their overall standard score. This led to significant confusion and frustration for the students.

After the first midterm, when students began demonstrating proficiency more readily (even on new material), the perception began to change. By the end of the semester, most students earned an A or B in the course, and we received many *strongly* positive comments about the grading scheme (including several students wishing that all their courses would adopt such a scheme).

## 4 SECOND ATTEMPT: STANDARDS-BASED GRADING (SUMMER & FALL 2021)

During Summer and Fall 2021, JAG and ML restructured the course to increase student buy-in. The two key changes involved reordering the content and restructuring the standards. The course structure was otherwise identical as in Spring and Fall 2020.

We begin by detailing changes we made to the standards.

**Proof by Induction**. We removed Loop Invariant Proofs from the course. We note that students already struggle with Proof by Induction, and it is a more transferable skill to other courses. Additionally, students found Loop Invariant Proofs to be unmotivated.

**Asymptotics.** We broke the Calculus standard into two separate standards: Calculus I Techniques (L'Hopital's Rule), and Calculus II Techniques (Ratio & Root Tests). As a result, our extensive problem set on this could be naturally divided into two smaller sections. This better conveyed to students their understanding of the respective techniques, while also reducing frustration at lack of partial credit.

**Spanning Trees**. We broke Prim's and Kruskal's algorithms into two standards, to build in early victories for students.

**Network Flows**. Due to the technical nature of residual graphs, we moved to working with the flow networks. We added a standard on Network Flow Terminology to ensure that students were comfortable reading a flow network and finding flow-augmenting paths. We also kept our Ford–Fulkerson standard, asking students to execute the algorithm on the flow network rather than using the residual graph construction. As a result, grading time for each of these standards was reported to be 20 seconds per problem, and over 90% of the class demonstrated proficiency twice for both these standards by the end of both the Summer and Fall 2021 offerings.

**Dynamic Programming.** We broke Dynamic Programming up into four key standards: Identifying subproblems, Writing down recurrence, Using the recurrence to fill in a lookup table, and Designing DP algorithms. The key techniques we stressed involved identifying the recursive structure (formalized with writing down a recurrence), and then using that recursive structure to construct and fill in an appropriate lookup table.

The final standard here, Designing DP algorithms, was only assessed twice via untimed problem sets. Students only had to demonstrate proficiency once in order to obtain credit for this standard. Because this standard was really a synthesis of several other standards, we felt that giving students time to think here was better for their learning, as opposed to testing this standard on a timed assessment.

**P vs. NP Problem.** We broke the P vs. NP standard up into four distinct standards: Formulating decision problems, Showing a problem belongs to P, Showing a problem belongs to NP, and Structure & Consequences of P vs. NP. This allowed us to better communicate key skills to students, as well as to use the Structure & Consequences of P vs. NP standard to challenge our students. Sample problems under the Structure & Consequences standard include asking students to recognize when someone is reducing in the wrong direction (a common mistake in NP-completeness proofs), the fact that the RSA cryptosystem can be broken if P = NP, and asking students to fill in the details of a proof showing that NP $\subseteq$ PSPACE.[2]

As a result, there were 30 content standards (plus the Engagement standard in Summer 2021). Each core topic also had a mix of mechanical standards (e.g., work through an algorithm, come up with a counter-example), as well as more involved synthesis and proofs-based standards. There were 23 mechanical standards, as well as the Engagement standard (more below). The cutoff for a C- was 23 standards in Summer 2021, and 20 standards in Fall 2021. In order to earn a B- or higher, it was necessary for students to demonstrate proficiency twice for a subset of the standards emphasizing synthesis or proofs-based problem solving.

**Reordering Content.** Aside from revising the standards, the key curricular change effectively involved moving greedy algorithms to the front of the course. In particular, we covered Proof by Induction, Pathfinding, Spanning Trees, and Network Flows (including Reductions to Max Flow) for the first midterm. Unlike in previous semesters, we did not receive complaints (let alone, organized complaints) from students as to the difficulty of the course. During the Summer 2021 session, several students who were repeating the course praised the change in the content ordering for helping the class to gain traction. At least one student reported to ML that the previous course ordering felt like "having a brick thrown in one's face."

The grades after the first midterm reflected the positive changes. In Summer 2021, the first quartile cutoff indicated that students had demonstrated proficiency twice for 6 standards and *at least* once for 10 standards. In Fall 2021, the first quartile cutoff indicated students had demonstrated proficiency twice for 6 standards and *at least* once for 9 standards. We contrast this with the Fall 2020 offering, where the first quartile cutoff students had demonstrated proficiency twice for 2 standards and *at least* once for 7 standards.

One concern in reordering the content is that it would make place the majority of the more difficult content (asymptotics, dynamic programming, and P vs. NP) in the second half of the course, possibly setting students up to fail. Despite these changes, we actually saw an increase in the number of students that passed the course (with a grade of at least C-)[3]. We had pass rates of 173/222 (77.93%) in Fall 2020, 51/64 (79.68%) in Summer 2021, and 199/253 (78.65%) in Fall 2021. We also note that in Fall 2020 and Fall 2021, we began the semester with approximately 270 students. So not only did our pass rate increase from Fall 2020 to Fall 2021, but the number of students who dropped also decreased.

[2]Complete writeups of these problems are available by request.

[3]We only count students who made it to the end of the course.

**Participation/Engagement Standard.** For Summer 2021, the Participation standard was renamed to be the Engagement standard. In order to earn credit for the Engagement standard, students had to complete a syllabus quiz and engage in at least 5/7 recitations. Each recitation was held in a remote synchronous format, with the TA facilitating active learning and group work. At the end of recitation, students submitted responses to a virtual exit ticket explaining in 1-2 sentences a concept they learned, as well as asking a question or identifying a confusing concept about the material from that day. Students who attended recitation and submitted a reasonably thoughtful response received credit for that week. Effectively, a thoughtful response required students to reflect, explain, or ask about a topic that was not definitional. For instance, *I learned about bipartite graphs* is not a thoughtful response, as a non-engaged student could simply look at the worksheet for that week. Instead, they would need to explain something they learned about bipartite graphs; for instance, *I learned that bipartite graphs have no cycles of odd length. Therefore all trees are bipartite.*

Students who attended recitation were generally engaged, though there were still some folks who did not discuss with their groups or ask questions. For students who did not wish to discuss in recitation, the exit ticket served as a means to assess their engagement. With rare exceptions, student responses on the exit tickets were quite thoughtful. The TAs also observed that students generally learned quite a bit in recitation, even if they struggled. There were 46/64 students who earned credit for the Engagement standard; of these, 41 (89.1%) students earned a passing grade. We contrast this with the pass rate of 51/64 (79.68%) students. While our goal was not to assess the benefits of active learning, it is worth noting that the gains amongst these students appear consistent with existing evidence on the effectiveness of active learning in increasing student learning and performance [6, 14, 28, 31].

Due to the logistics of having in-person and remote recitations in Fall 2021, we did not include a Participation or Engagement standard.

## 5 DISCUSSION AND CONCLUSION

**What Worked Well.** Our SBG scheme had several key advantages. When grading, the course staff were able to focus on providing feedback rather than determining appropriate quantities of partial credit. This allowed us to grade more quickly. Additionally, in instances where multiple graders were assigned to the same problem, we observed increased consistency in grading compared to previous semesters [4].

Standards-based grading also enabled us to clearly communicate to students their understanding of the concepts and techniques. This reduced student anxiety around grades, in that students could readily identify specific standards on which to focus, based on their desired end of semester grades. Furthermore, SBG allowed students to focus on growth without being hampered down by low grades resulting from earlier assignments that fell short of expectations. In

[4]Though grading consistency remains an issue at this scale.

addition, students were less likely to dispute their grades; and those that did dispute their grades were less apt to negotiate for points on individual problems. Conversations in office hours centered around the course content rather than individual student standing.

**Online Quizzes.** At the onset of the COVID-19 pandemic, we migrated our quizzes to Canvas. We kept this format in subsequent semesters, in part due to the class being held in either a remote or hybrid format. The online format of the quizzes affords a number of advantages that make it worth permanently retaining. First, the instructional staff are responsible for providing accommodations to students with 1.5x testing time. For in-person quizzes, we must schedule a time to proctor these students, which is a logistical difficulty. Students who miss class also often want to make up the in-person quizzes, which creates an additional logistical burden. Using Canvas allowed us to easily grant individual students extra testing time or extended deadlines, which was a significant convenience.

**Reassessment Tokens.** The reassessment mechanism greatly contributed to student learning and reducing student anxiety. Student exposition on the first round of corrections & reflections often demonstrated a significantly stronger understanding of the content than the original problem set or quiz submissions. For students whose first correction & reflection submissions needed additional work, the additional discussion with the instructional staff served to largely clarify remaining misconceptions. The revised correction & reflection submissions reflected this. The overwhelming majority of students demonstrated proficiency on the reassessment quizzes. In cases where students had extenuating circumstances, we granted free additional retake tokens. This allowed students to focus on their circumstances in the short-term and attempt the material at a later point, when they were able to better devote the time.

Developing the reassessment quizzes- particularly for our more involved standards like Exchange Arguments and Reductions to Max Flow- and processing student corrections & reflections resulted in a significant undertaking. We received close to 380 submissions during the semester, with roughly 225 of those in the last 24 hours. We allocated two TAs to process these, so that students could take their reassessment quizzes and have grades back within 24 hours of the final opening. As most students wait until the last 24 hours to submit their corrections & reflections, we recommend opening it roughly six weeks before the end of the semester and closing it at least 1.5 weeks before the final. This minimizes the need to create multiple reassessment quizzes for a given standard, which is particularly helpful for the more involved standards.

**Participation/Engagement Standard.** The Participation standard incentivized students to attend recitations, though without necessarily engaging with any active learning activities. The exit ticket requirement introduced in Summer 2021 successfully incentivized engagement, which correlated with a higher pass rate. While tracking exit tickets was feasible in a smaller summer class with remote synchronous recitations, it would not have been in a larger Fall or Spring offering. For these reasons, we have discontinued the Participation/Engagement standard.

**Instructional Workload.** The two key challenges centered around developing course infrastructure (problem sets, quizzes, solutions), as well as the additional grading load incurred by the weekly quizzes. This was particularly problematic in Spring 2020, when we developed the course from scratch. In subsequent semesters, we reused existing infrastructure, though a significant amount was still developed from scratch in Fall 2020 and Summer 2021. We recommend developing the problem set and quizzes for a given standard at once. Our TAs also spent almost all of their allotted hours on grading, which left little time for developing course infrastructure and other logistical tasks. The situation greatly improved once we hired graders, though the hiring process took several weeks. The grading situation improved in subsequent semesters, as we requested and received two additional graders.

**Student Performance.** We have two CS degrees: one by admissions (BS), and one open enrollment (BA).[5]. Algorithms has always been required for the BS degree and became required for the BA degree in Fall '19. Thus, our results can most closely be compared to the Fall '19 semester.

From Fall '15-'19, Algorithms witnessed decreased performance commensurate with growing enrollments. This was especially drastic for the BA CS population. We conjecture that this is partly due to difficulties faced in the department's Discrete Math (DM) course, such as enrollment sizes, multiple learning outcomes[6], and no TA support. These barriers may have disproportionately left the BA CS students without sufficient support to thrive when they reach Algorithms. Without TAs in DM, it would be a Herculean task to implement known effective strategies like active learning [6, 14, 28, 31], frequent low-stakes assessment [7, 10, 19, 20], and detailed feedback [8, 17, 26, 27].

As our class population is significantly different between Spring and Fall semesters[7], we report pass rates[8] for the Fall semester. Our data suggests students in Fall iterations of Algorithms perform notably better under SBG. We had pass rates of 178/222 (80.18%) in Fall '17 and 165/202 (81.68%) in Fall '18. In Fall '19, when Algorithms became required for the BA degree, the pass rate dropped to 163/231 (70.56%)[9]. Under our SBG scheme, we improved the pass rate to 173/222 (77.93%) in Fall '20. Under the changes in Summer and Fall '21 (see Sec. 4), the pass rates grew slightly to 51/64 (79.68%) and 199/253 (78.65%), respectively. The initial enrollments were comparable (~270-280 students) in Fall '19, '20, and '21, and so the changes discussed in Sec. 4 correlated with more students passing and fewer students dropping the course in Fall '21.[10]

## ACKNOWLEDGMENTS

We thank our course staff for their hard work, S. Krinsky for sharing her course structure and advice during our initial planning, V. Dunn for compiling detailed grade data, and the anonymous referees for helpful feedback. JAG thanks C. Heckmann, B. Hayes, D. Larremore, B. Waggoner, and A. Clauset for useful feedback. ML thanks B.H. Payne for useful feedback, A. Book and A. Edwards for helpful discussions regarding the changes made in Summer '21, and A. Book for helpful feedback on the development of a course text [11]. JAG was partially supported by NSF CAREER award CCF-2047756.

[5]The BS degree is housed in the College of Engineering, and the BA degree is housed in the College of Arts & Sciences

[6]Content, synthesis-based problem solving, mathematical formalisms

[7]Notably in terms of majors, year in program, number of students taking the class a second time

[8]Again, we consider students who completed the course.

[9]ML worked with the course directly in Fall '19 & '20, and believes that without SBG, the pass rate in Fall '20 would have been similar to Fall '19.

[10]Due to space constraints, see [3] for full discussion.

## REFERENCES


[1] Ian D. Beatty. 2013. Standards-based grading in introductory university physics. *Journal of the Scholarship of Teaching and Learning* 13 (2013), 1–22. ERIC number EJ1011679.

[2] Tom Buckmiller, Randal E. Peters, and Jerrid Kruse. 2017. Questioning Points and Percentages: Standards-Based Grading (SBG) in Higher Education. *College Teaching* 65 (2017), 151 – 157. https://doi.org/10.1080/87567555.2017.1302919

[3] Lijun Chen, Joshua A. Grochow, Ryan Layer, and Michael Levet. 2022. Experience Report: Standards-Based Grading at Scale. (04 2022). https://arxiv.org/abs/2204.12046 Extended Version of ITiCSE 2022 Conference Paper.

[4] Philip Duker, Anna Gawboy, Bryn Hughes, and Kris P. Shaffer. 2015. Hacking the Music Theory Classroom: Standards-Based Grading, Just-in-Time Teaching, and the Inverted Class. *Music Theory Online* 21 (2015). https://doi.org/10.30535/mto.21.1.2

[5] Jason Elsinger and Drew G. Lewis. 2019. Applying a Standards-Based Grading Framework Across Lower Level Mathematics Courses. *PRIMUS* (2019), 1–23. https://doi.org/10.1080/10511970.2019.1674430

[6] Scott Freeman, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111, 23 (2014), 8410–8415. https://doi.org/10.1073/pnas.1319030111

[7] Ufuk Guven. 2017. *The Relationship Between Testing Frequency and Student Achievment in Eight-Grade Mathematics: An International Comparative Study Based on TIMSS 2011.* Ph. D. Dissertation. https://doi.org/10.13140/RG.2.2.32786.71363

[8] John Hattie and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77, 1 (2007), 81–112. https://doi.org/10.3102/003465430298487

[9] Silvia Heubach and Sharona Krinsky. 2020. Implementing Mastery-Based Grading at Scale in Introductory Statistics. *PRIMUS* (2020), 1–17. https://doi.org/10.1080/10511970.2019.1700576

[10] Jeffrey D. Karpicke and Janell R. Blunt. 2011. Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science* 331, 6018 (2011), 772–775. https://doi.org/10.1126/science.1199327

[11] Michael Levet. 2021. CSCI 3104 Algorithms- Lecture Notes. https://michaellevet.github.io/Algorithms_Notes.pdf

[12] Drew D Lewis. 2020. Gender Effects on Re-assessment Attempts in a Standards-Based Grading Implementation. *PRIMUS* 30 (2020), 539–551. https://doi.org/10.1080/10511970.2019.1616636

[13] Robert J. Marzano. 2011. *Formative Assessment & Standards-Based Grading.* Solution Tree.

[14] Jeffrey J. McConnell. 1996. Active Learning and Its Use in Computer Science. In *Proceedings of the 1st Conference on Integrating Technology into Computer Science Education* (Barcelona, Spain) *(ITiCSE '96).* Association for Computing Machinery, New York, NY, USA, 52–54. https://doi.org/10.1145/237466.237526

[15] Kate Owens. 2015. A Beginner's Guide to Standards Based Grading. https://blogs.ams.org/matheducation/2015/11/20/a-beginners-guide-to-standards-based-grading/

[16] Scott L. Post. 2017. Standards-Based Grading in a Thermodynamics Course. *iJEP* 7 (2017), 173–181. https://doi.org/10.3991/ijep.v7i1.6472

[17] Renah Razzaq, Korinn S. Ostrow, and Neil T. Heffernan. 2020. Effect of Immediate Feedback on Math Achievement at the High School Level. In *Artificial Intelligence in Education*, Ig Ibert Bittencourt, Mutlu Cukurova, Kasia Muldner, Rose Luckin, and Eva Millán (Eds.). Springer International Publishing, Cham, 263–267. https://doi.org/10.1007/978-3-030-52240-7_48

[18] Douglas B Reeves. 2008. *Making standards work.* Lead + Learn Press.

[19] Henry L. Roediger and Andrew C. Butler. 2011. The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences* 15, 1 (2011), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

[20] Henry L. Roediger III, Adam L. Putnam, and Megan A. Smith. 2011. Chapter One - Ten Benefits of Testing and Their Applications to Educational Practice. Psychology of Learning and Motivation, Vol. 55. Academic Press, 1–36. https://doi.org/10.1016/B978-0-12-387691-1.00001-6

[21] Matt Salomone. 2018. Standards-Based Grading: My Implementation. http://matthematics.com/standards-based-grading-my-implementation/

[22] Matt Salomone. 2018. Standards-Based Grading: Origins. http://matthematics.com/standards-based-grading-origins/

[23] Megan E. Selbach-Allen, Sarah J. Greenwald, Amy Ksir, and Jill E. Thomley. 2020. Raising the Bar with Standards-Based Grading. *PRIMUS* (2020), 1–17. https://doi.org/10.1080/10511970.2019.1695237

[24] Michael Shindler, Matt Ferland, Aaron Cote, and Olivera Grujic. 2020. Experience Report: Preemptive Final Exams for Computer Science Theory Classes. *J. Comput. Sci. Coll.* 35, 10 (April 2020), 9–14. https://dl.acm.org/doi/abs/10.5555/3417699.3417700

[25] Katherine E. Stange. 2018. Standards-based Grading in an Introduction to Abstract Mathematics Course. *PRIMUS* 28, 9 (2018), 797–820. https://doi.org/10.1080/10511970.2017.1408044

[26] Roar Bakken Stovner and Kirsti Klette. 2022. Teacher feedback on procedural skills, conceptual understanding, and mathematical practices: A video study in lower secondary mathematics classrooms. *Teaching and Teacher Education* 110 (2022), 103593. https://doi.org/10.1016/j.tate.2021.103593

[27] Roar B. Stovner, Kirsti Klette, and Guri A. Nortvedt. 2021. The instructional situations in which mathematics teachers provide substantive feedback. *Educational Studies in Mathematics* 108 (2021), 533–551. Issue 3. https://doi.org/10.1007/s10649-021-10065-w

[28] Elli J. Theobald, Mariah J. Hill, Elisa Tran, Sweta Agrawal, E. Nicole Arroyo, Shawn Behling, Nyasha Chambwe, Dianne Laboy Cintrón, Jacob D. Cooper, Gideon Dunster, Jared A. Grummer, Kelly Hennessey, Jennifer Hsiao, Nicole Iranon, Leonard Jones, Hannah Jordt, Marlowe Keller, Melissa E. Lacey, Caitlin E. Littlefield, Alexander Lowe, Shannon Newman, Vera Okolo, Savannah Olroyd, Brandon R. Peecook, Sarah B. Pickett, David L. Slager, Itzue W. Caviedes-Solis, Kathryn E. Stanchak, Vasudha Sundaravardan, Camila Valdebenito, Claire R. Williams, Kaitlin Zinsli, and Scott Freeman. 2020. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences* 117, 12 (2020), 6476–6483. https://doi.org/10.1073/pnas.1916903117

[29] Matt Townsley. 2019. Walking the talk: Embedding standards-based grading in an educational leadership course. *Journal of Research Initiatives* 4 (2019), 2. https://doi.org/10.32674/jsard.v6i2.3419

[30] Cathy Vatterott. 2015. *Rethinking grading: meaningful assessment for standards-based learning.* ASCD.

[31] Janice D. Yoder and Catherine M. Hochevar. 2005. Encouraging Active Learning Can Improve Students' Performance on Examinations. *Teaching of Psychology* 32, 2 (2005), 91–95. https://doi.org/10.1207/s15328023top32022