Specifications Grading at Scale: Improved Letter Grades and Grading-Related Interactions in a Course with over 1,000 Students

Kate J. McKnelly,‡ William J. Howitz,‡ Taylor A. Thane, & Renée D. Link*

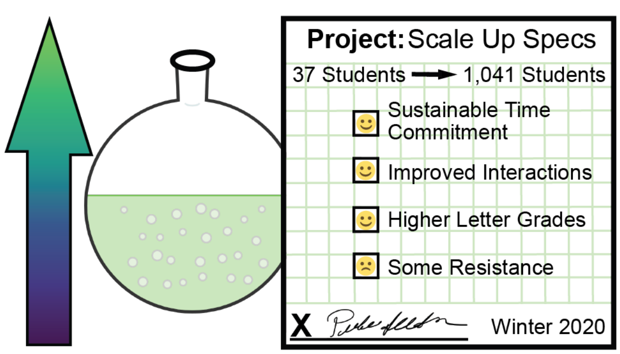
Department of Chemistry, University of California—Irvine,

Irvine, CA 92697, United States

ABSTRACT

In our previous work we piloted a specifications grading system in an organic chemistry laboratory course with 37 students. Our current work describes the scale up of that specifications grading system to a course with over 1,000 students. Strategies used for keeping the system manageable and mitigating the time commitment required to do so are described. We found that the time necessary to grade student work and manage the specifications grading implementation of the course was not any greater than for the previous, points-based course, that grade-related interactions were more positive, and that student letter grades increased. Despite the increase in final letter grades, we encountered some resistance to the grading system from students and graduate teaching assistants. Here, we explore their concerns and address the difficulty of alternative grading methods in overcoming habituation to traditional points-based grading systems. Future work is needed to evaluate student and GTA buy-in, to assess potential improvement in student work, and to address questions regarding equity in specifications grading systems.

GRAPHICAL ABSTRACT



KEYWORDS
Alternative Grading, Second-Year Undergraduate, Graduate Teaching Assistants, Organic Chemistry Laboratory Instruction, Collaborative/Cooperative Learning, Testing/Assessment, Resistance to Change

INTRODUCTION

The logistics of grading courses for which enrollment nears or exceeds 1,000 students are challenging. This is particularly true for laboratory courses where, for example, 50 or more laboratory sections, at 20 students per section, are required to accommodate the 1,000 plus student enrollment. Having a large number of laboratory sections necessitates the employment of a large number of graduate student teaching assistants (GTAs), which adds additional challenges to grading logistics. To mitigate grading challenges associated with running a large laboratory course at the University of California, Irvine (UCI), we previously piloted a smaller organic chemistry laboratory course using a scalable specifications grading system with plans to subsequently transition our large enrollment organic chemistry laboratory courses to specifications grading systems.[1]

Specifications grading was popularized by Linda Nilson in 2014.[2] Under this grading system, students earn their desired course grade by satisfactorily completing set bundles or modules of assignments. The instructor defines the assignments that comprise each bundle or module so course learning outcomes (LOs) are directly aligned with course grades. Student assignments are commonly assessed as satisfactory or unsatisfactory using rubrics in which meeting set numbers of criteria define passing thresholds, eliminating the need for partial credit. If students do not earn a satisfactory assessment, they may choose to revise and resubmit their work to be reassessed. Nilson recommends that the number of opportunities for revision and resubmission be limited to keep instructor workload manageable. The number of times students may revise and resubmit assignments is commonly limited using a token system whereby students may exchange their limited number of tokens for resubmission.[1,3–17]

Examples of specifications grading systems implemented in college-level courses began appearing in peer-reviewed journals and in conference proceedings in 2016, and the number published each year continues to grow (Table 1). While published examples in journals, conference proceedings, and conference abstracts include courses from nearly all disciplines,[18–21] including humanities,[15,22,23] social sciences,[4,24–26] education,[9,27] public administration,[28] health sciences,[29–31] and veterinary medicine,[16,32] STEM courses comprise the majority. Examples of specifications grading systems in STEM courses come from engineering,[5,10,12,33–41] computer science,[42–49] mathematics,[3,7,8,11,50–52] biological sciences,[7,53–55] physics,[7,56] and chemistry.[1,6,7,14,17,57–61] In the chemistry discipline, our pilot study is the only publication that specifically describes implementing specifications grading in a standalone laboratory course.[1]

Table 1. Examples Of STEM Courses That Have Implemented Specifications Grading Systems Published in Peer-Reviewed Journals or As Conference Papers.

Discipline	Course	Number of Students	Year
Chemistry	Organic Chemistry[57]	35	2017
Chemistry	General Chemistry[58]	32	2017
Physics	Acoustics (Lecture and Laboratory)[56]	NA[b]	2017
Mechanical Engineering	Kinematics[34]	28	2017
Mechanical Engineering	Thermodynamics[37,38,d]	18 (each course)	2018
Chemistry	General Chemistry[59]	60-75	2018
Mathematics	Foundations of Mathematics[3]	10-20	2018
Computer Science	Computer Architecture[44]	21	2018
Computer Science	Requirements Engineering[45]	NA[b]	2018
Civil Engineering	Structural Design in Reinforced Concrete[36]	≤35	2018
Chemistry	General Chemistry[6]	24-35	2019
Chemistry, Physics, Biology, Mathematics	Principles of Chemistry I and II (and 10 others)[7]	≤30	2019
Mechanical Engineering	Statics/Dynamics[40]	NA[b]	2019
Mechanical Engineering	Heat and Mass Transfer[39,d]	15	2019
Biomedical Engineering	Bioelectricity[5]	17	2019
Mathematics	Matrix Algebra and Systems of Differential Equations, Discrete and Combinatorial Algebra[8]	20-25	2020
Mathematics	Fundamentals of Algebra and Calculus[50,d]	NA[b]	2020

Mathematics	Mathematics for Preservice Elementary School Teachers Course[11,a]	NA[b]	2020
Chemistry	Organic Chemistry II[60,c]	12	2020
Chemical Engineering	Senior Laboratory Course and Undergraduate Research Course[13,a]	31, 15-30	2020
Mathematics	Ordinary Differential Equations[51]	≤26	2020
Electrical and Computer Engineering	Linear Circuits Analysis Laboratory 1[12]	31	2020
Engineering	Senior Capstone Design, Biomedical Engineering Statistics, First-Year Engineering Course[10,a]	26, 93, 218	2020
Chemistry	Organic Chemistry Laboratory[1]	37	2021
Biology	Cell Biology[55]	24	2021
Chemistry	Writing for Chemists[14]	NA[b]	2021
Computer Science	Intro Programming for Web Applications, Data Structures, Algorithms, Cybersecurity, Systems[43]	≤32	2021
Computer Science	Game Programming Course[42,a,e]	NA[b]	2021
Chemistry	Organic Chemistry[61]	65-85	2021
Engineering	Engineering Capstone Course[10,a]	193 (across six courses)	2021
Mathematics	Discrete Math[52]	128	2022
Chemistry	Analytical Chemistry[17]	NA[b]	2022

[a]Course name not specified
[b]NA = not specified in the paper
[c]Course began in person and converted to online during COVID19
[d]Hybrid course delivery
[e]Online course delivery

In our pilot study, we designed and implemented a specifications grading system in the smallest of the three organic chemistry laboratory courses at UCI.[1] Similar to many of the published examples of courses using specifications grading systems, we cited transparency, student agency, and a greater focus on learning as motivations for adopting the new grading system. Bundling course assignments and implementing a satisfactory/unsatisfactory assessment system provide transparent ties between the final course grades students earn and their successful achievement of course LOs. Clear requirements for each course grade facilitate greater student agency, allowing students to choose where to focus their time and effort in the course based on their target grade. Removing partial credit shifts the focus of student-instructor interactions from earning points to understanding course content. The shift away from partial credit could also reduce grading time, but this reduction could be offset by the need to grade resubmissions — a concern relevant to one raised by Ring about the feasibility and management of grading with a specifications grading system in larger courses.[57] In this paper we address the concern about the feasibility of a scaled up specifications grading system by reporting our adaptation of the system from our pilot course that accommodates 200-student and 1,000-student scale organic chemistry laboratory courses.

SCALING A SPECIFICATIONS GRADING SYSTEM FOR A LARGE LABORATORY COURSE

We chose to implement our specifications grading system, designed for scalability, in a mid-sized 200-student fall term course and a large 1,000-student winter term course. At UCI the large organic chemistry laboratory courses are taken by all students who need the course except for chemistry majors and honors students. These courses consist of a weekly, 50 minute laboratory lecture taught by the instructor and a weekly 3 hour and 50 minute laboratory section facilitated by a GTA. During the regular academic year, the entire organic chemistry laboratory course sequence described here is coordinated and taught by a single faculty member in a teaching-track position with the equivalent of tenure (R.D.L.).[62,63] The laboratory courses are offset from the organic chemistry lecture courses, so students on the normal course sequence complete their organic chemistry laboratory courses in the winter and spring terms (Figure 1). The full organic chemistry laboratory course sequence comprises three courses, but most students only need to complete the first two courses for their degree requirements. The winter and spring on-sequence courses have enrollments that average over 1,000 students. Because these courses are so large, there is a dedicated Head TA to support the instructor. The Head TA is responsible for assisting with the instruction of the laboratory lecture classes, running the weekly staff meeting, hosting office hours, and printing and distributing the laboratory practical exams. Students taking the laboratory courses off-sequence take the first course during the summer term and the second course during the fall term. The fall course has an average of about 200 students. The students in the mid-sized fall term and large winter term courses experienced our specifications grading system for the first time.

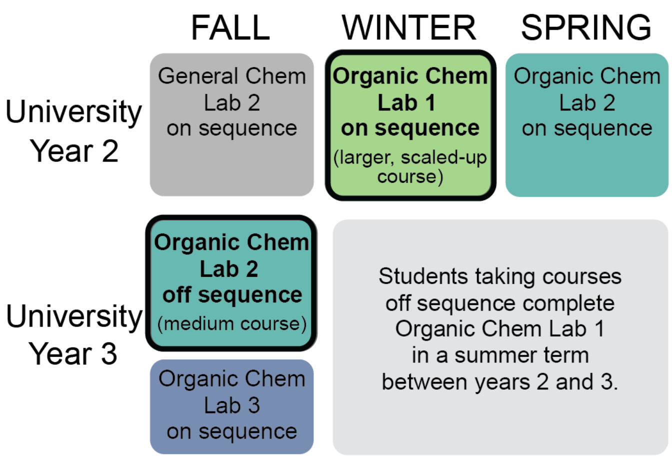


Figure 1. Sequence of organic chemistry laboratory courses at the University of California, Irvine.

In the 208-student fall course and 1,041-student winter course, we leveraged the existing grading structure developed for the pilot course (Table 2). Similar assignment bundles were crafted for letter grades and students could track their progress toward earning their desired letter grade using a provided grade tracker. The instructor adapted a computer script using the R statistical programming language to permit the calculation of final letter grades in an expedient manner. Grading rubrics for notebook pages had previously been designed by type of experiment (technique-focused, reaction-focused, etc.) rather than written for individual experiments, which made them suitable for use in all of the organic chemistry laboratory courses. Although the post-laboratory assignment rubrics had to be modified for the content covered in the fall and winter term courses, rubric criteria language was reused when possible. For example, we used the same criteria for the conclusion section of each postlab assignment. Revision and resubmission of post-laboratory assignments were controlled using the same token system used in the pilot course and token trade-in requests were monitored using a Google form embedded in the course learning management system (LMS). An example token trade-in list and token use workflow are provided in the Supporting Information. To reduce the burden on GTAs to regrade revised student assignments, students were only required to revise the section(s) of the assignment they did not earn credit for on the rubric and highlight the changes they made so they could be found easily. More exam versions were needed in the larger courses, but we limited the number needed by giving the same exam version to all laboratory sections that met at the same time each week.

Table 2. Logistical Differences Between Pilot Course and Larger, Scaled-Up, Course.

	Pilot Course	**Mid-Size**	**Large**
Place in Course Sequence	3rd (off sequence)	2nd (off sequence)	1st (on sequence)
Term	Summer	Fall	Winter
Students Enrolled	37	208	1,041
TAs for Course	2	5	34
Laboratory Lecture Sections	1	2	4
Individual Laboratory Sections	2[a]	10	68
Exam Versions Needed			
Knowledge Check	3	5	10
Safety	6	6	6
Technique	3[b]	3[b]	3[b]
Mastery	4	6	12

[a]Each TA teaches 2 lab sections per week during the summer term, so the number of sections TAs teach is different than during the normal academic year terms.
[b]There are three versions of exam documents provided, but the quantities of reagents for the practical are varied and thus result in more exam variations.

To help students understand the specifications grading system and prepare for the exams, we considered what we learned from the pilot course and applied those considerations to the fall and winter courses. To prevent student misconceptions about how final course grades are determined (Table 3), we established student buy-in by reviewing the student grade tracker in detail on the first day of the class meetings. We also reviewed the tracker with GTAs during the first weekly staff meeting. Students were reminded about how to determine their final course grade at various points throughout the course in the laboratory lecture and through announcements on the LMS because we anticipated there would still be some student and GTA confusion. To allay student fears about the "all-or-nothing" grading aspect of the specifications grading system, we also began introducing rubric items and overall assignment grades as "satisfactory" or "needs revision" rather than "satisfactory" or "unsatisfactory." We reemphasized that students could use tokens to resubmit assignments if they did not achieve a satisfactory assessment. To facilitate student practice on critical thinking questions that would be seen on the mastery portion of the practical exam, the section most students struggled with in the pilot course, we incorporated example problems during the weekly 50-minute laboratory lecture class, posted extra practice problems to the course LMS, and provided solutions to those practice problems with explicit satisfactory criteria listed.

Table 3. Letter Grade Requirements For The Large, Scaled-Up Course Using The Specifications Grading System.

	Criteria from Specifications Grading System	
Course Requirements	Course Grade Level[a,b]	Set of Criteria Completed
Online Pre-laboratory Homework Assignments	A	90 - 100 % correct
	B	80 - 100 % correct
	C	70 - 100% correct
	D	< 70% correct
Pre-laboratory Video Quizzes	A	85 - 100 % correct
	B	85 - 100 % correct
	C	75 - 100% correct
	D	< 75% correct
Laboratory Notebook Assignments	A	7 Satisfactory
	B	6 - 7 Satisfactory
	C	5 - 6 Satisfactory
	D	4 Satisfactory
Post-laboratory Assignments	A	5 Satisfactory + 1 full written laboratory report
	B	4 Satisfactory
	C	3 Satisfactory
	D	2 Satisfactory
Laboratory Lecture Participation	A	7 required
	B	6 required
	C	4 - 5 required
	D	< 4 required
Practical Exam	A	Passed Mastery Final Passed Knowledge Check w/S Passed 3 Lab Techniques Passed 4/6 safety questions
	B	Passed Mastery Final Passed Knowledge Check w/S

	Passed 2 Lab Techniques Passed 4/6 safety questions
C	Passed Knowledge Check w/S Passed 1 Lab Technique Passed 4/6 safety questions
D	< above criteria

[a]Students who do not meet the minimum criteria for D grade earn an F in the course.
[b]UCI does include plus and minus grades, and our course does have set criteria for students to achieve +/- grades. We have left them out here for clarity. The Supporting Information contains full criteria for all letter grades.

LARGE COURSE IMPLEMENTATION OUTCOMES

To demonstrate that specifications grading is indeed scalable to laboratory courses with at least 1,000 students, and to share lessons learned, we report instructor, Head TA, GTA, and student perceptions of the specifications grading system and changes in student grades. These outcomes are organized in sections focused on how much time was necessary to manage the course and grade assignments, how student interactions with the instructor, Head TA, GTAs, and other students changed, and how letter grades shifted from the points-based course to the specifications grading course. To assess GTA and student perceptions, we administered surveys to both the points-based and specifications grading courses. We surveyed GTAs (specifications grading course n = 16, 55% response rate; points-based course n = 19, 58% response rate) to determine if student questions were more frequently about point attainment or comprehension of course content. We asked students for their thoughts about the grading system in general, what worked well, and what should be changed. Because responses from students in both the medium and large courses were similar, only the large course results (n = 794, 76% response rate) are discussed. We end by presenting comparisons of student letter grades and the quality of student work on one assignment in courses with and without a specifications grading system. The paper focuses on the comparison between the pilot and larger winter term course, but we include some data from the mid-sized fall term course as a reference for instructors who teach courses of that scale.

Time to Manage and Implement a Course Using a Specifications Grading System

A prominent concern voiced by some faculty about implementing a specifications grading system in a large course is the potential for unsustainable time demands to be placed on the instructor. These demands include the time needed to design and implement the course under the new grading system, the time needed to manage the token system, and the time needed to grade revised student work.

Time to Manage Course. We found that the time to manage the course was not a concern. Both the instructor (R. D. L.) and the Head TA (T. A. T.) found that the time spent on managing course logistics throughout the term did not change between the specifications grading course and the prior points-based course. However, we found that the initial development of the specifications grading system did require additional time, as is inevitably expected with any new course development. The bulk of the development time was dedicated to revising existing rubrics and designing additional versions of practical exams.

Time to Manage Token System. We found that managing the token system in the specifications grading course did not require an increase in overall workload. In the points-based course, each student was allocated a single due date extension and a single make-up opportunity. To request either opportunity, students had to email the instructor. With the implementation of a token system, we returned agency to students. Our system provided additional opportunities to use tokens beyond the single due date extension and make-up opportunity and gave students the choice of how to allocate their tokens in the ways that best fit their individual needs. Although time was needed to manage the token system, this time was offset by the marked decrease in the time the instructor and Head TA spent responding to emails from students requesting due date extensions or other course policy exceptions. Because no reason was needed for a student to use a token as a late pass or to make up a missed laboratory, the token system also removed the affective labor required for the instructor or Head TA to judge whether or not a student's reason for requesting an assignment extension or other course policy exception was for a "valid" reason. Students experiencing major extenuating circumstances were still encouraged to contact the instructor or Head TA to coordinate necessary, broader accommodations.

Time to Grade Assignments. Contrary to common concerns, the instructor, Head TA, and GTAs did not report spending increased time grading assignments in the specifications grading course as compared to time spent in the points-based course.

The instructor and Head TA were responsible for facilitating and grading the multiple-choice knowledge check exams. Passing the knowledge check exam was required for students to pass the course, so an exam retake was offered for students who did not pass on the first attempt. Additional knowledge check exam versions were prepared from a pool of multiple-choice questions in advance of administering the first exam, so there was not a significant time requirement for the instructor and Head TA to administer exam retakes. In both the fall and winter courses, fewer than ten percent of the students needed to take the knowledge check exam a second time to meet the passing threshold. Because we used a multiple-choice format, additional time required to grade exam retakes was minimal.

GTAs were responsible for grading revised laboratory notebook assignments and post-laboratory reports. To assess the amount of time GTAs spent grading these assignments, we asked our GTAs — in both the points-based and specifications grading versions of the course — to report the number of hours they spent grading them (Q1 and Q2, Table 4). The median number of hours GTAs reported spending on these assignments were the same in the specifications grading course and the points-based course, with five hours spent on laboratory notebook assignments and seven hours spent on post-laboratory reports. While the mean values did increase in the specifications grading course, this was the result of a GTA who spent an unusually large amount of time on grading (see Supporting Information). These results suggest that there is not an increased grading burden in the specifications grading course as compared to prior iterations of the course using a traditional points-based grading system.

Table 4. Self-reported Hours GTAs Spent Grading Course Assignments. (Specs n = 16, points n = 19)

Survey Question	Type	Mean (SD)	Median (IQR)
Q1. On average, how long did it take you each week to grade notebook (pre-lab) pages?	Specs	5.19 (0.83)	5 (1)
	Points	4.89 (0.46)	5 (0)
Q2. On average, how long did it take you each week to grade post-labs?	Specs	7.38 (0.96)	7 (1)
	Points	6.95 (0.91)	7 (1.5)

Although GTAs in the specifications grading course reported spending the equivalent amount of time grading as GTAs in the points-based course, in a post-course survey, they expressed concerns about the time they spent grading assignments based on the quality of the initial work submitted by students. Because students were able to revise and resubmit post-laboratory reports assessed as needing revision, some GTAs felt students might be turning in lower quality work for their first submissions to buy additional time to work on assignments after the due dates. This led some GTAs to feel that they were spending more time grading in the specifications grading course than they would have otherwise if they were in a points-based course. A suggested approach to mitigate this concern was to require students to spend more tokens to revise and resubmit an assignment that was assessed as needing revision. The GTA that recommended this approach felt that it would incentivize students to try harder and produce higher quality work on the first submission of an assignment, reducing the time needed to be spent on grading revised work. Despite GTA perceptions that they spent more time to grade and regrade assignments in the specifications grading course than in the points-based course, the total time GTAs reported grading was not greater, even with the added load of grading the revised assignments.

Shifts in Interactions

Although the time needed to administer the course and to grade student work in the specifications grading course did not change as compared to the points-based course, we observed three main types of interaction changes between the points-based course and the specifications grading course. Student comments from surveys and anecdotal reports from the instructor, Head TA, and GTAs suggested that student–student interactions shifted, as many students felt their learning environment with their peers was markedly more collaborative and less competitive. Additionally, when interacting with students, the instructor and Head TA observed a marked shift in student questions from requesting additional points on assignments to understanding their misconceptions about course content — one of the reported benefits of specifications grading.[2] Finally, interactions between the instructor and students concerning the final exam changed from a focus on how students' final letter grades were determined to why student work was or was not satisfactory and the impact of the final exam on their final letter grade.

In the specifications grading course, the instructor and Head TA noticed that students collaborated more often in office hours settings. Students similarly recognized that their grade was no longer dependent on others and that there was greater incentive to collaborate because there was no longer a course curve. Students self-reported their reflections on the collaborative environment in the specifications grading course in the post-course survey:

> "Overall, I think the grading system has an emotional positive impact for students by removing the competition elements of the class."
> "Our own grade is determined by us entirely. We are not blindly hoping for a grade and unfortunately depending on others doing worse."
> "Because the class is not curved, it seems that students are little [*sic*] more relaxed and willing to help each other. It also encourages students who may not usually strive for a high grade to do so because students are not pitted against each other."

In addition to changes in student–student interactions, the instructor and Head TA observed changes in the way students interacted with them during office hours, through email, and in passing interactions. The students who attended the Head TA's office hours in the specifications grading course appeared less anxious than students who attended in the points-based course. They appeared to be focused on learning from their mistakes to better understand the course material, being motivated to do so when they needed to revise assignments that had not earned a satisfactory assessment. This apparent growth mindset was not observed as frequently in the points-based system where students appeared to be more motivated to come to office hours to argue for partial credit on graded assignments rather than to understand their mistakes.

Survey responses from GTAs supported the anecdotal reports from the instructor and Head TA about changes in their interactions with students. The median number of student emails GTAs received regarding disagreements about grading on assignments dropped from five in the points-based course to four and a half in the specifications grading course, and the median number of in-person interactions on the same topic dropped from five in the points-based course to four in the specifications grading course (Q1 and Q3, Table 5). In contrast, the median number of emails GTAs received from students seeking assistance to better understand the course material increased to six in the specifications grading course, from five in the points-based course (Q2, Table 5). The median number of in-person interactions remained the same in the two courses (Q4, Table 5). Although the changes were not statistically significant,[64–66] the survey results are consistent with the anecdotal reports from the instructor and Head TA.

Table 5. GTA Responses to Survey Questions Regarding Quantity of Student Emails and Interactions. (Specs n = 16, points n = 19)

Survey Question	Type	Mean (SD)	Median (IQR)
Q1. How many emails did you receive about a student disagreeing about whether they met the minimum criteria for a satisfactory score on rubric items in general? (Points version: How many emails did you receive about "points" in general?)	Specs	5.06 (1.48)	4.5 (1.25)
	Points	4.84 (1.01)	5 (1)
Q2. How many emails did you receive about student understanding in general?	Specs	6.38 (1.59)	6 (2.5)
	Points	5.21 (1.47)	5 (2)
Q3. How many interactions did you have with students in-person about students disagreeing about whether they met the minimum criteria for a satisfactory score on rubric items in general? (Points version: How many interactions did you have with students in-person about "points"?)	Specs	4.44 (0.63)	4 (1)
	Points	4.58 (0.61)	5 (1)
Q4. How many interactions did you have with students in-person about student understanding?	Specs	6.50 (1.83)	6 (2.5)
	Points	6.21 (1.75)	6 (2.5)

Another interaction between the instructor and the students that changed was the way the students communicated with the instructor about final exams and letter grades. Discussions shifted from a focus on students not understanding how their assignment and exam scores translated to their letter grade to a focus on questions about why work on exams did not meet the passing threshold for a desired letter grade and concerns about the impact one exam could have on the final letter grade. This shift may have occurred because the specifications grading system eliminated the need for a normalization process, and the transparency of the system made clear to students the impact of poor exam performance on their final letter grade.

In the points-based course, students often contacted the instructor at the end of the term to ask how their assignment and exam scores correlated to their letter grade, to voice concerns about a student in a different section earning a higher letter grade after normalization despite having lower scores on some assessments, to ask for their grade to be rounded up, or to argue for more partial credit on assignments. The students' questions made it evident that they did not understand the impact of practical exams on their final grade because they did not fully understand how their letter grade was determined using grade normalization. The necessity of normalizing section averages due to variations in GTA grading and exam versions caused another set of difficult conversations with students because students in sections with lowered averages after normalization felt that their final grade was unfair. Both of these types of interactions likely arose because the process of determining their final letter grade was not transparent.

Under the specifications grading system, the end-of-term conversations between students and the instructor centered around questions about how students could improve their work to meet passing thresholds on the mastery exam and concerns about student performance on a single exam determining their letter grades. When contacting the instructor about their exam performance specifically, students asked more conceptual understanding questions, such as why their responses to the questions on the mastery component of the practical exam did not meet expectations for credit. Despite the transparency of the specifications grading system, some students went into the practical exam expecting A-level final course grades based on their performance on other assessments in the course. After performing poorly on one or more components of the practical final exam, some students earned B-level or C-level grades, but they still felt that they deserved an A-level grade in the course. The student perception that poor performance on a single assessment should not affect their final letter grade also occurred in the points-based course, but the instructor received even more emails regarding this perception in the specifications grading course, likely because the impact of performing poorly on the final exam was clearer. Without a curve to raise their final grade at the end, students knew immediately what the impact of poor performance on the exam was and contacted the instructor to voice their concerns.

Additional GTA and Student Perceptions and Suggestions

In addition to the beneficial changes in student–student, student–GTA, and student–instructor interactions after transitioning to the specifications grading course, students and GTAs found value in other aspects of the new grading system. However, we also encountered some resistance to the specifications grading system in student and GTA feedback. This resistance, in addition to the previously mentioned student concern about the impact of the final exam on their final grade, occurred despite multiple efforts to establish buy-in for the specifications grading system.

GTAs and students found several aspects of the specifications grading system — opportunities for revision, the token system, the removal of the competitive environment, and increased grading transparency — valuable. The GTAs found the binary assignment grading to be more clear, organized, and efficient, and they liked the ease with which they felt it facilitated consistent grading. A number of GTAs reported that the ample opportunities for students to improve their assignments through revision when needed, by way of the token system, seemed to reduce their students' stress because they did not have to earn a satisfactory assessment on each assignment to earn a "good" grade.

Students included comments on what they liked about the specifications grading system in their responses to open-ended survey questions (see Supporting Information for question prompts), and these positive responses were consistent with the benefits cited by the GTAs. Some representative student comments are included after this paragraph. Students appreciated having the option to correct their mistakes to earn credit and improve their grade. Some students also commented that the option to correct their mistakes encouraged them to revisit their work when they would not have done so in a points-based course, leading to an increased focus on learning and understanding the material. As discussed previously, students found the grading system to be more transparent and equitable because they were no longer competing with one another for letter grades. Students also felt the flexibility offered by the token system (e.g., opportunities to revise and resubmit assignments and submit assignments late) lowered their stress and anxiety about the course.

> "I think getting the students to actually revise the works gets the student to go back and try to understand what went wrong. It helps them see what went wrong and not only can they improve their grade, they also see how they can learn from their mistakes."
>
> "I was able to fix what I got wrong. In other courses I would check my grade and not care about the mistakes I made. This way I was able to go back and take the time to revise for credit."
>
> "I think the concept of tokens was very helpful especially when midterm season came along. It allowed for us to get all the things we need to done and not have to worry about prioritizing a class over another."
>
> "I liked that we had tokens to earn and use in case of emergencies or other unexpected situations."

While GTAs and students appreciated many aspects of the specifications grading system, both groups suggested changes, some of which indicated misunderstandings about the new grading system or a desire to return to a more familiar system. Some GTAs felt that the specifications grading system penalized students who only missed the threshold for a satisfactory assessment on an assignment by one or two rubric items, which gave students the same outcome on an assignment as a student who did not submit any work. GTAs desired rubrics that were either more specific or that gave them opportunities to reward student effort through partial credit. The GTA suggestions included offering a half-credit option or reverting back to a points-based rubric. Another recommendation was to lower the threshold for earning a satisfactory assessment. Students who were one rubric item away would then earn credit for the assignment. Consequently, GTAs would not have to spend as much time discussing with students whether their initial work met the passing threshold.

Students suggested similar changes to the specifications grading system as those suggested by GTAs, some quotes of which follow this paragraph. Similar to feelings expressed by some GTAs, some students felt they should be rewarded with partial credit for the effort put into the assignments and that the thresholds for passing assignments should be lowered. The recommendations for lowering the passing thresholds commonly came up in cases where students were one or two rubric items below the threshold so reducing the threshold meant they would not need to spend a token to revise and resubmit the assignment.

> "There should be some partial credit given back for effort/things done right"
>
> "For example the all or nothing approach can be frustrating sometimes especially if I was off by only a little bit"
>
> "lower some thresholds"

These feelings from the students and GTAs are reflected in observations made by the Head TA about how students discussed their token usage during office hours. Most students managed their tokens well, but the Head TA reflected that there were some students who mismanaged their tokens in one of two ways. Some students failed to appropriately ration their tokens initially (e.g. spending them mostly on late passes early in the course), leaving them with fewer opportunities to revise and resubmit assignments later in the academic term. Other students were too hesitant to use their tokens earlier in the quarter to revise and resubmit assignments, leaving them with fewer overall opportunities to pass the number of assignments necessary to earn their letter grade.

Letter Grades in Specifications Grading System vs. Points-Based System

Despite some students' concerns about their course grade in the specifications grading course, especially in regards to the impact of the final exam and lack of partial credit, their actual grades were higher than their student counterparts in the points-based course. In the large course, a comparison of final letter grades in the specifications grading course (n = 1,041) to those from a points-based course (n = 1,189) — taught by the same instructor the year prior — showed that in the specifications grading course, there was an increase in the total percentages of A-level and B-level grades (57% in the points-based course, 87% in the specifications grading course, Figure 2).[64,65,67] The total percentages of C-level and D-level grades decreased correspondingly (41% in the points-based course, 11.5% in the specifications grading course, Figure 2), and the amount of failing grades (3%) did not change. A similar trend was observed in the mid-sized course with an even larger decrease in total percentages of C-level and D-level grades (61% in the points-based course, 18% in the specifications grading course, Figure 2).

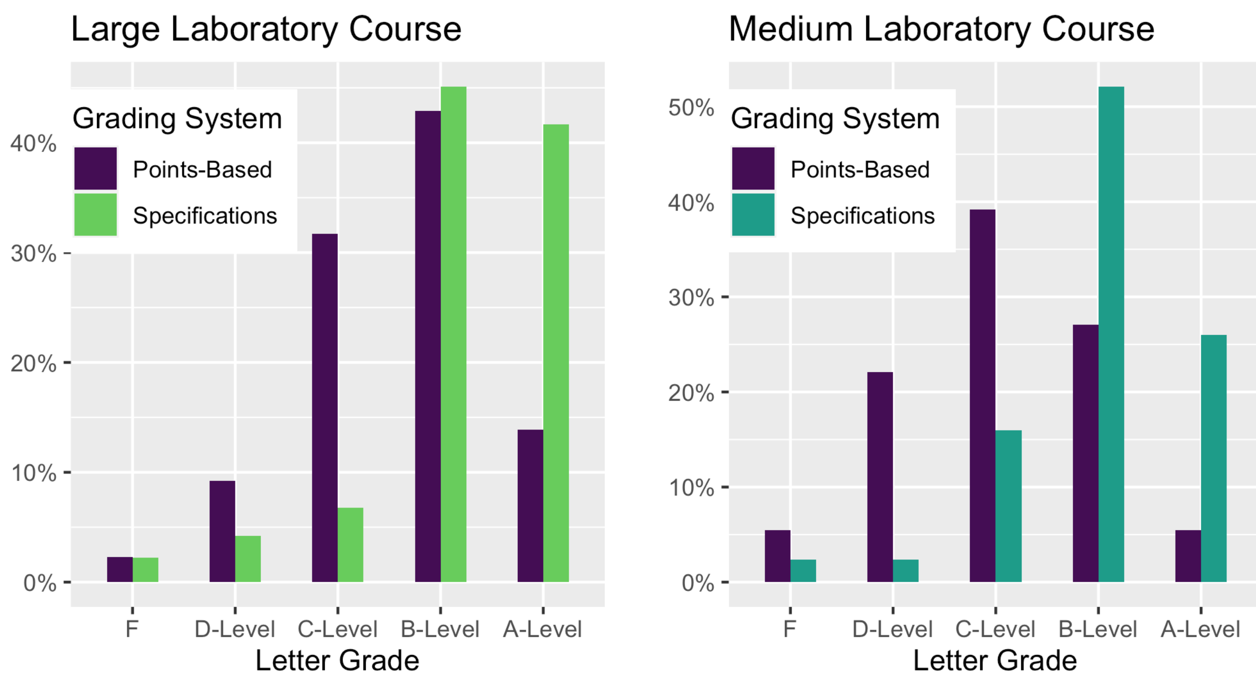


Figure 2. Grade distributions of the course using a points-based system (previous course) and the current course with the specifications grading system. See supporting information for the distribution including +/- grades.

Why Did Grades Go Up?

The shift to higher letter grades that we observed under the specifications grading system is not unique to the courses reported here. We observed similar increases in our pilot course.[1] In other specifications grading STEM courses where authors have tracked changes in final letter grades, similar increases in A-level and/or B-level grades and commensurate decreases in C-level grades were observed.[5,8,15,36,55,58,59] While some authors did not track changes in final letter grades, they observed an increase in the quality of student work or an improvement in student performance on exams in their specifications grading courses as compared to their prior points-based courses.[6,12,55,68] For example, Ring compared students' final exam performance in both courses by regrading work from students in the points-based course using the rubric from the specifications grading course, which did not allow for partial credit.[68] When grading final exams without allocating any partial credit, Ring found that students in the specifications grading course performed better on the final exam than students in the points-based course.

To determine if the higher grades earned in our large specifications grading course, compared to the previous points-based course, were the result of higher quality work, we compared the student submissions of one post-laboratory report assignment in each of the two courses. We randomly selected a total of 60 submissions, 30 student post-laboratory reports from each course. Reports from the third experiment in the course were selected because that experiment occurred toward the middle of the academic term after students had adjusted to the specifications grading system. We chose to grade all reports with the points-based rubric rather than the specifications grading rubric because it was less detailed than the specifications grading rubric; we did not want to bias the results in favor of the specifications grading system and potentially cause lower scores for student reports from the points-based course.

We recruited six graders from GTAs who had taught the course in at least one prior term and assigned each grader an equal number of student reports from the points-based grading course and the specifications grading course. Each student report was graded by two different graders, and scores from the two graders were averaged. To minimize potential bias, the graders were not informed about which reports were from students in the points-based course and which were from students in the specifications grading course.

The results of the grading comparison indicate that there was not a significant difference in the quality of student laboratory reports between the points-based grading course and the specifications grading course (Figure 3). The median scores for the reports are both 25 out of 40 points. Although we did not find any statistically significant differences in the quality of a small sample of a single assignment, a limitation of the small sample size is that we could not assess the quality of work across the entire student population or the quality of all submitted assignments in the course. It is possible that the quality of student work may have improved for some assignments and not others.

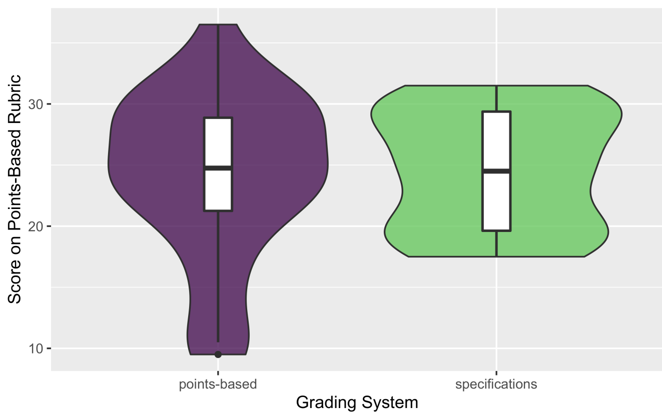


Figure 3. Grading comparison of student acid-base extraction laboratory reports submitted in a points-based grading course (n = 30) and the specifications grading course (n = 30). All reports were graded with a points-based rubric because the specifications grading rubric added additional criteria, which would render grading of reports completed with a points-based rubric a poor comparison.

An additional explanation for the higher letter grades observed in the large specifications grading course could be the result of removing the course curve used in the previous points-based courses. Avoiding a curve in the specifications grading course and determining final letter grades based solely on achievement of the course LOs may better reflect the true grades students should be earning, implying student grades earned in previous points-based courses may have been artificially and unnecessarily lowered.

Assigning final letter grades using curves is based on flawed assumptions and practices. A curve creates a competitive environment in which higher letter grades are artificially treated as scarce resources that students earn by outperforming their peers.[69] In addition, implementing a curve is based on the assumption that student performance will follow a normal distribution without consideration of the external cultural and social pressures that contribute to letter grade scores.[70] A normal distribution should not necessarily arise when comparing student work, even if the student population in a given course is large.[71,72] Assigning grades based on a curve is a ranking of student performance relative to one another rather than a measure of how well students met the course LOs. Students at the lower end of a curve based on a normal distribution are assigned final letter grades that are the result of comparisons to their peers instead of reflections of their achievement of the LOs.[73–75]

Trying to compare letter grades in courses that use a curve to those that use specifications grading is challenging, and is akin to comparing apples and oranges, because a curve is a ranking and specifications grading is a measurement of how well students met course LOs. Because students earn letter grades in specifications grading systems based on whether or not they have met the LOs for a course, their letter grades should better reflect their learning rather than their rank in comparison to their peers. If students have met the LOs, they perform well in the course. We are not concerned about grade inflation in our course because we are confident in the LOs we set and the work required of students to meet those LOs.

LARGE COURSE IMPLEMENTATION REFLECTIONS AND CONCLUSIONS

We found that scaling the specifications grading system from the pilot study course (37 students)[1] to a mid-sized course (208 students) and to a large course (1,041 students) and managing the courses during the academic term was feasible and successful. Concerns about the time requirements placed on the instructor when adopting a specifications grading system are mentioned in the literature[57,62,76] and, in our experience, are brought up frequently in discussions at seminars, conference presentations, and workshops. However, we found that the time to manage the course throughout the term did not exceed that of the points-based course. Others who have implemented specifications grading similarly found the time to design the grading system was substantial,[4,15,20,28,29,36,51,52,77] but that the time to grade assessments and manage the course did not considerably change or decreased.[1,3,4,8,9,14,15,17,19,28,36,41,43,44,52,58] The only major time commitment we encountered was the design of the grading system up front, but it is our experience that courses with preexisting student-centered structures and policies will require less time to convert to specifications grading.

One of the motivations for adopting specifications grading is increased student agency over their learning, and indeed, we saw a positive shift in student interactions with the instructor, Head TA, GTAs, and their peers that suggested they were more focused on their learning. Compared to the points-based course, the specifications grading course had a larger number of As and Bs. However, there were no significant changes in the quality of the small subset of student reports we examined between a course iteration that used a points-based system and one that used a specifications grading system.

Despite the increase in As and Bs, similar to the pilot course, we encountered some resistance from students and GTAs to the grading system. Some students expressed anxiety about the perceived high stakes of the finals in which poor performance on one assessment could bring their letter grade down. Some students (and some GTAs) also felt the lack of partial credit and the instructor-defined passing thresholds were disadvantages that negatively impacted student performance in the course. Neither of the concerns expressed by students and GTAs above should dissuade others from adopting specifications grading. We take the next several paragraphs to provide context for these concerns and address them.

While the student perception of the high stakes of the practical exams is understandable, we have not found a way of removing this aspect of the specifications grading system. The practical exam components are the only assessments students complete individually in a controlled setting, and we feel these assessments are necessary for students to demonstrate their ability to meet the course LOs independently without help from peers. However, we took steps to mitigate the perceived stressful nature of these assessments by allowing retakes in the case of the knowledge check exam, keeping the A-level and B-level thresholds reasonable — 50% and 25%, respectively — for the mastery exam, and allowing students to use tokens in exchange for a small increase to their mastery exam score when they were close to the A-level or B-level threshold. In the case of students who completed all A-level or B-level work but did not score well enough on the mastery exam to earn their target grade even when using tokens to increase their score, we instituted a B- safety net to ensure that these student's grades would be no lower than a B-. Additionally, if a student did not pass the knowledge check exam after the second attempt, we implemented a grade negotiation. During the negotiation, students are asked to propose what they believe is an appropriate grade based on their full body of work and to consider the fact that they were not able to pass an exam after two attempts. They were also required to reflect on why they were not able to pass the exam and describe what they can do to avoid such situations in the future. In nearly all of these cases students responded with a reasonable proposed grade, comparable to what the instructor would have assigned, and they reflected on ways they could improve their study approaches and time management skills or work on managing test anxiety. In a few cases students were far too hard on themselves, and the instructor responded with a higher grade and a suggestion to be kinder to themselves.

The habituation of students to points-based systems in which students are extrinsically motivated to argue for partial credit points rather than to revise work to earn credit is likely a major contributing reason for some of their observed resistance to specifications grading. It has been well-documented that the extrinsic motivation of getting a desired grade "crowds out" the benefits of a student's intrinsic motivation to learn.[78–81] In traditional points-based grading systems this extrinsic motivation leads to gamesmanship where students focus more on the accumulation of points rather than the learning process.[82] Because specifications grading requires students to complete specific bundles of assignments at instructor-defined passing thresholds rather than to earn a select number of points across all assignments in the course, a primary source of students' extrinsic motivation has been modified. Other faculty who have adopted specifications grading have reported similar student resistance associated with their lack of familiarity with the grading system.[1,4,5,10,20,34,35,43,62] Resistance to other alternative grading systems with similar characteristics to specifications grading, including contract grading and standards-based grading, is also attributed to the habituation of students to points-based grading systems.[83,84]

Another possible reason that students are resistant to having to revise assignments rather than the potential to earn partial credit points may stem from their perception that needing to revise work is evidence of failure. We attempted to mitigate this perception in the medium and large courses discussed in this paper by changing the language of attaining rubric criteria to "satisfactory" or "needs revision". However, this change in language may not have been sufficient. In our course, and potentially others that use specifications grading systems, students may hold that earning a "needs revision" assessment is equivalent to failure, an idea that stems from the self-worth theory of academic achievement.[85,86] The theory proposes that self-worth that is based on academic performance leads to helplessness and anxiety in the face of failure. A study of student perceptions of standards-based grading, which is similar in many ways to specifications grading, found that some students viewed reassessment as an indication of lower intelligence, which had consequences for social status.[84]

Because of the social value some students placed on intelligence, there was a potential desire to avoid reassessment.

The habituation to points-based systems also explains why some students and GTAs made arguments to reincorporate partial credit. Partial credit is often viewed by students as a reward for effort rather than for the quality of their work, and previous studies have suggested that students believe about 40% of their grade should reflect the effort put into the course.[87,88] A study by Greenberger found that 34.1% of students believed they should earn at least a B for completing the course readings and 40.7% for attending all classes.[89] The belief that effort should have such a large contribution to the final letter grade is often discordant with faculty-evaluated performance.[90] The fact that specifications grading systems allot no reward for effort on assessments that do not meet a passing threshold may further exacerbate this discordance between how students and faculty regard how, if at all, effort should be weighted in assessment.

The desire of some GTAs in our course to reward effort using partial credit is also not surprising given that GTAs who are resistant to innovative teaching methods prefer to rely on their own experiences as students to inform how they teach.[91] Additionally, Seymour found that GTAs who resisted innovative teaching methods resented having to grade in the same manner as other GTAs, viewing this requirement as an infringement on their professional prerogative as teachers.[91] Evidence from a study by Yerushalmi et al. corroborates these findings, noting that GTAs desired rubrics with flexibility built in so they could have greater discretion in assigning scores. Some GTAs indicated they would grade based on intuition, inferring student understanding from their work, rather than using the provided rubrics. Other GTAs were inclined to grade more leniently because they expressed being able to relate to and identify with their students' experiences, having been former undergraduate students themselves.[92] This last observation is further supported by work which found that GTAs' perceptions and interpretations of their prior experiences influenced their instructional practices.[93,94]

By the time students enter college and GTAs enter graduate school, most have been habituated and indoctrinated by years spent in courses with points-based systems. It is not surprising that they exhibit resistance to grading systems that they are not familiar and comfortable with. Specifications grading is not widely used, so students in our course likely encountered it for the first time. Extra care and effort must be taken in courses using a specifications grading system to establish buy-in for both students and GTAs. The instructor should be as transparent as possible and indicate the similarities and differences between the specifications grading system and points-based systems. Experiences described in the literature about pedagogical innovations such as active learning initiatives suggest that student and GTA resistance to specifications grading will likely diminish with repeated exposure.[91,95,96]

Specifications grading systems are still new, and there are many avenues left to investigate and explore. While we did not identify significant differences in the quality of student performance on one assignment in our study, a more nuanced comparison and evaluation of student work in points-based courses versus specifications grading courses would yield insight into whether or not specifications grading systems facilitate increased student learning. This evaluation provides a ripe opportunity for instructors who, after the thick of the pandemic, have not yet transitioned their courses to specifications grading. In the realm of buy-in, investigating how students and GTAs buy-in to specifications grading over time and whether students view the need to revise work as akin to failure could yield insight into how and when instructors should implement buy-in interventions during a course and could help identify the characteristics of students and TAs who exhibit the most habituated resistance. Finally, there are many questions left still to ask about if and how specifications grading systems improve equity in the classroom, such as whether the increase we have observed in letter grades diminishes systemic grade disparities experienced by students from marginalized groups and how students are benefitting from specifications grading systems in the affective domain.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available:

Supporting Information includes an IRB statement, Student Grade Tracker, Token Trade-In List, Flow Chart to Manage Token Economy, TA and Student Open-Ended Questions, and Full Letter Grades Graph Including +/-Grades. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: rlink@uci.edu

Author Contributions

‡K.J.M and W.J.H. contributed equally.

Notes

The authors declare no competing financial interests. K.J.M.'s current institutional affiliation is Emory University. W.J.H.'s current institutional affiliation is Georgia Institute of Technology. T.A.T.'s current institutional affiliation is California State Polytechnic University, Pomona.

ACKNOWLEDGMENTS

We thank Casadora Boone, Katia Hatem, Katelyn Haduong, Selina Sanchez, and Emma Wilfong for their feedback on rubric design. Simon Lam provided integral support in course implementation. We also thank Amanda Chung, Bryant Lim, Sarah Wang, and Taylor Frey for their help with the post-laboratory report grading comparison. Kameryn Denaro was instrumental in aiding us in the design of the grading comparison and advising us on the statistical analysis. Finally, we thank the UCI Chemistry Department for support and Chem 51L students and GTAs for participating.

REFERENCES

(1) Howitz, W. J.; McKnelly, K. J.; Link, R. D. Developing and Implementing a Specifications Grading System in an Organic Chemistry Laboratory Course. *J. Chem. Educ.* **2021**, *98* (2), 385–394.

(2) Nilson, L. B.; Stanny, C. J. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*, Reprint edition.; Stylus Publishing: Sterling, VA, 2014.

(3) Williams, K. Specifications-Based Grading in an Introduction to Proofs Course. *Problems, Resources, and Issues in Mathematics Undergraduate Studies* **2018**, *28* (2), 128–142.

(4) Blackstone, B.; Oldmixon, E. Specifications Grading in Political Science. *Journal of Political Science Education* **2019**, *15* (2), 191–205.

(5) Brian P. Helmke. Specifications Grading in an Upper-Level BME Elective Course. In *ASEE Annual Conference & Exposition*; ASEE Conferences, 2019. https://doi.org/10.18260/1-2--33278.

(6) Martin, L. J. Introducing Components of Specifications Grading to a General Chemistry I Course. In *Enhancing Retention in Introductory Chemistry Courses: Teaching Practices and Assessments*; ACS Symposium Series; American Chemical Society, 2019; Vol. 1330, pp 105–119.

(7) Tsoi, M. Y.; Anzovino, M. E.; Erickson, A. H.; Forringer, E. R.; Henary, E. Variations in Implementation of Specifications Grading in STEM Courses. *Georgia Journal of Science* **2019**, *77* (2), 0_1,0_2,1–21.

(8) Carlisle, S. Simple Specifications Grading. *Problems, Resources, and Issues in Mathematics Undergraduate Studies* **2020**, *30* (8-10), 926–951.

(9) Dennen, V.; Bagdy, L. Shifting to Specifications Grading: Two Design Cases. In *TCC 2020 Conference Papers*; 2020.

(10) Fernandez, T.; Martin, K.; Mangum, R.; Bell-Huff, C. Whose Grade Is It Anyway?: Transitioning Engineering Courses to an Evidence-Based Specifications Grading System. In *2020 ASEE Virtual Annual Conference Content Access Proceedings*; ASEE Conferences, 2020. https://doi.org/10.18260/1-2--35512.

(11) Prasad, P. V. Using Revision and Specifications Grading to Develop Students' Mathematical Habits of Mind. *PRIMUS* **2020**, 1–18.

(12) Cosoroaba, E. Helping Students Write It Right: Instilling Good Report-Writing Habits in a Linear Circuit Lab Course. In *2020 ASEE Virtual Annual Conference Content Access*. https://doi.org/10.18260/1-2--34723.

(13) Pascal, J.; Vogel, T. J.; Wagstrom, K. Grading by Competency and Specifications: Giving Better Feedback and Saving Time.

(14) McKnelly, K. J.; Morris, M. A.; Mang, S. A. Redesigning a "Writing for Chemists" Course Using Specifications Grading. *J. Chem. Educ.* **2021**, *98* (4), 1201–1207.

(15) Vitale, S. E.; Concepción, D. W. Improving Student Learning with Aspects of Specifications Grading. *Teach. Philos.* **2021**, *44* (1), 29–57.

(16) Hofmeister, E. H.; Fogelberg, K.; Conner, B. J.; Gibbons, P. Specifications Grading in a Cardiovascular Systems Course: Student and Course Coordinator Perspectives on the Impacts on Student Achievement. *J. Vet. Med. Educ.* **2022**. https://doi.org/10.3138/jvme-2021-0115.

(17) Hunter, R. A.; Pompano, R. R.; Tuchler, M. F. Alternative Assessment of Active Learning. In *Active Learning in the Analytical Chemistry Curriculum*; ACS Symposium Series; American Chemical Society, 2022; Vol. 1409, pp 269–295.

(18) Bonner, M. W. Grading Rigor in Counselor Education: A Specifications Grading Framework. *Educational Research Quarterly* **2016**, *39* (4), 21.

(19) Elkins, D. M. Grading to Learn: An Analysis of the Importance and Application of Specifications Grading in a Communication Course. *Kentucky Journal of Communication* **2016**, *35* (2), 2016.

(20) LeHew, M. Specifications Grading: Experimenting with a New Form of Student Assessment. *Journal for Academic Excellence* **2019**, *6* (4), 12–15.

(21) Walden, P. R. Student Motivation, Anxiety and Pass/Fail Grading: A SoTL Project. *Teaching and Learning in Communication Sciences & Disorders* **2022**, *6* (1), 13.

(22) Blodgett, B. J. Grading Matters in Theological Education. *Teach Theol Relig* **2017**, *20* (4), 314–326.

(23) Earl, D. Two Years of Specifications Grading in Philosophy. *Teach. Philos.* **2021**. https://doi.org/10.5840/teachphil20211118154.

(24) Shields, K.; Denlinger, K.; Webb, M. Not Missing the Point (s): Applying Specifications Grading to Credit-Bearing Information Literacy Classes. In *The Grounded Instruction Librarian: Participating in The Scholarship of Teaching and Learning*; Association of College and Research Libraries: Chicago, IL, 2019.

(25) Rivera, J.; Groleau, T. Student and Faculty Transformations from Teaching Wicked Geography Problems: A Journey of Transdisciplinary Teaching between Business and Geography. *Journal of Geography in Higher Education* **2021**, *45* (4), 538–548.

(26) Wasniewski, E.; Munro, T.; Tandon, T. Exploring Online Specifications Grading: An Undergraduate Course Case Study. In *EdMedia + Innovate Learning*; Bastiaens, T. J., Ed.; Association for the Advancement of Computing in Education (AACE), 2021; pp 179–184.

(27) Quintana, R.; Quintana, C. When Classroom Interactions Have to Go Online: The Move to Specifications Grading in a Project-Based Design Course. *Information and Learning Sciences* **2020**, *121* (7/8), 525–532.

(28) Jones, P. A. Implementing Specifications Grading in MPA Courses: A Potential Strategy for Better Work-Life Balance. *Journal of Public Affairs Education* **2020**, 1–17.

(29) Ganesh, C.; Smith, J. Using Multiple High-Impact Practices to Reduce Bottlenecks and Improve Student Learning in an Undergraduate Health Science Program. *Journal of the Scholarship of Teaching and Learning* **2017**, *17* (2), 74–84.

(30) Leslie, P.; Lundblom, E. Specifications Grading: What It Is, and Lessons Learned. *Semin. Speech Lang.* **2020**, *41* (4), 298–309.

(31) Pope, L.; Parker, H. B.; Ultsch, S. Assessment of Specifications Grading in an Undergraduate Dietetics Course. *J. Nutr. Educ. Behav.* **2020**, *52* (4), 439–446.

(32) Hofmeister, E. H.; Fogelberg, K.; Conner, B. J.; Gibbons, P. Specifications Grading for Veterinary Medicine. *J. Vet. Med. Educ.* **2022**. https://doi.org/10.3138/jvme-2021-0116.

(33) Bell-Huff, C. L.; Fernandez, T. M.; Morgan, K. L.; LeDoux, J. M. WIP: How Students Externalize Epistemologies: Describing How Students Explain, Ground, and Consciously Construct Their Definitions of Engineering and Biomedical Engineering. In *2019 ASEE Annual Conference & Exposition.* https://doi.org/10.18260/1-2--33563.

(34) Mirth, J. A. Applying Lean Thinking to the Structure and Delivery of a Kinematics Course. In *2017 ASEE Annual Conference & Exposition.* https://doi.org/10.18260/1-2--27603.

(35) Martin, K. M.; Fernandez, T. M.; Mangum, R. Work in Progress: Examining Engineering Seniors Students' Perception of Justice and Fairness of Grading Practices. In *2021 ASEE Virtual Annual Conference Content Access*; ASEE Conferences: Virtual Conference, 2021.

(36) Matthew D. Lovell P.E. Board 44: Defining and Assessing Competencies in an Undergraduate Reinforced Concrete Design Course. In *2018 ASEE Annual Conference & Exposition*; ASEE Conferences, 2018. https://doi.org/10.18260/1-2--30034.

(37) Mendez, J. Standards-Based Specifications Grading in a Hybrid Course. In *2018 ASEE Annual Conference & Exposition*; 2018.

(38) Mendez, J. Standards-Based Specifications Grading in Thermodynamics. In *2018 ASEE Annual Conference & Exposition*; 2018.

(39) Mendez, J. Development of a Hybrid Heat and Mass Transfer Course. In *ASME 2018 International Mechanical Engineering Congress and Exposition*; American Society of Mechanical Engineers Digital Collection, 2019. https://doi.org/10.1115/IMECE2018-87934.

(40) Mirth, J. A. A Specifications-Based Approach for the Design and Delivery of a Statics/Dynamics Course. In *2019 ASEE Annual Conference & Exposition*; ASEE Conferences, 2019.

(41) Jennifer Pascal; Troy J. Vogel; Kristina Wagstrom. Grading by Competency and Specifications: Giving Better Feedback and Saving Time. In *2020 ASEE Virtual Annual Conference Content Access*; ASEE Conferences, 2020. https://doi.org/10.18260/1-2--34712.

(42) Gestwicki, P. Godot Engine and Checklist-Based Specifications: Revising a Game Programming Class for Asynchronous Online Teaching. *J. Comput. Sci. Coll.* **2021**, *37* (4), 30–40.

(43) Sanft, K. R.; Drawert, B.; Whitley, A. Modified Specifications Grading in Computer Science: Preliminary Assessment and Experience across Five Undergraduate Courses. *J. Comput. Sci. Coll.* **2021**, *36* (5), 34–46.

(44) Mirsky, G. M. Effectiveness of Specifications Grading in Teaching Technical Writing to Computer Science Students. *J. Comput. Sci. Coll.* **2018**, *34* (1), 104–110.

(45) Rupakheti, C. R.; Hays, M.; Mohan, S.; Chenoweth, S.; Stouder, A. On a Pursuit for Perfecting an Undergraduate Requirements Engineering Course. *J. Syst. Softw.* **2018**, *144*, 366–381.

(46) Roberson, C. Techniques for Using Specifications Grading in Computer Science. *J. Comput. Sci. Coll.* **2018**, *33* (6), 192–193.

(47) Myers, D. S. Designing Specifications Grading Systems. *J. Comput. Sci. Coll.* **2021**, *37* (5), 91–92.

(48) McGuffee, J. Syllabus Transformation for Standards-Based and Specifications Grading in CS Courses. *J. Comput. Sci. Coll.* **2018**, *33* (5), 54.

(49) Largent, D. My Exploration of Specifications Grading in a Discussion-Based Course. *J. Comput. Sci. Coll.* **2017**, *33* (1), 89.

(50) Gratwick, R.; Kinnear, G.; Wood, A. K. An Online Course Promoting Wider Access to University Mathematics. *Proceedings of the British Society for Research into Learning Mathematics* **2020**, *40* (1).

(51) Henriksen, M.; Kotas, J.; Wentworth, M. Specifications-Based Grading Reduces Anxiety for Students of Ordinary Differential Equations. *CODEE Journal* **2020**, *13* (1), 1.

(52) Tuson, E.; Hickey, T. 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*; ITiCSE ’22; Association for Computing Machinery: New York, NY, USA, 2022; pp 19–25.

(53) Rylaarsdam, R. P.; Heinz, C. Specifications Based Grading: Nearly Pointless Education That Points Students to Key Concepts. *The FASEB Journal* **2016**, *30* (S1), 662.13–662.13.

(54) Sosa Ponce, M. L.; Moorhead, G. B. G. Developing Scientific Writing Skills in Upper Level Biochemistry Students through Extensive Practice and Feedback. *The FASEB Journal* **2020**, *34* (S1), 1–1.

(55) Katzman Shoshana D.; Hurst-Kennedy Jennifer; Barrera Alessandra; Talley Jennell; Javazon Elisabeth; Diaz Mary; Anzovino Mary E. The Effect of Specifications Grading on Students’ Learning and Attitudes in an Undergraduate-Level Cell Biology Course. *J. Microbiol. Biol. Educ.* **2021**, *22* (3), e00200–e00221.

(56) Ludwigsen, D. Effective and Efficient Grading: Tips for Assessing Student Learning. *Proc. Mtgs. Acoust.* **2017**, *30* (1), 025001.

(57) Ring, J. ConfChem Conference on Select 2016 BCCE Presentations: Specifications Grading in the Flipped Organic Classroom. *J. Chem. Educ.* **2017**, *94* (12), 2005–2006.

(58) Toledo, S.; Dubas, J. M. A Learner-Centered Grading Method Focused on Reaching Proficiency with Course Learning Outcomes. *J. Chem. Educ.* **2017**, *94* (8), 1043–1050.

(59) Hollinsed, W. C. Applying Innovations in Teaching to General Chemistry. In *Increasing Retention of Under-Represented Students in STEM through Affective and Cognitive Interventions*; ACS Symposium Series; American Chemical Society, 2018; Vol. 1301, pp 145–152.

(60) Houseknecht, J. B.; Bates, L. K. Transition to Remote Instruction Using Hybrid Just-in-Time Teaching, Collaborative Learning, and Specifications Grading for Organic Chemistry 2. *J. Chem. Educ.* **2020**, *97* (9), 3230–3234.

(61) Ahlberg, L. Organic Chemistry Core Competencies: Helping Students Engage Using Specifications. In *Engaging Students in Organic Chemistry*; ACS Symposium Series; American Chemical Society, 2021; Vol. 1378, pp 25–36.

(62) Streifer, A. C.; Palmer, M. S. Is Specifications Grading Right for Me?: A Readiness Assessment to Help Instructors Decide. *College Teaching* **2021**, 1–8.

(63) Supiano, B. Can a Teaching Track Improve Undergraduate Education? *The Chronicle of Higher Education*. The Chronicle of Higher Education July 19, 2022.

(64) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.

(65) Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; Kuhn, M.; Pedersen, T.; Miller, E.; Bache, S.; Müller, K.; Ooms, J.; Robinson, D.; Seidel, D.; Spinu, V.; Takahashi, K.; Vaughan, D.; Wilke, C.; Woo, K.; Yutani, H. Welcome to the Tidyverse. *JOSS* **2019**, *4* (43), 1686.

(66) de la Rubia, E. A.; Zhu, H.; Ellis, S.; Quinn, M.; McNamara, A. *skimr: Compact and Flexible Summaries of Data*. https://CRAN.R-project.org/package=skimr (accessed 2020-11-12).

(67) Garnier, S. *viridis: Default Color Maps from "matplotlib."* https://CRAN.R-project.org/package=viridis (accessed 2020-11-12).

(68) Ring, J. Specifications Grading in the Flipped Organic Classroom; 2016.

(69) Holland, D. G. The Struggle to Belong and Thrive. In *Talking about Leaving Revisited*; Seymour, E., Hunter, A.-B., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp 277–327.

(70) Bowen, R. S.; Cooper, M. M. Grading on a Curve as a Systemic Issue of Equity in Chemistry Education. *J. Chem. Educ.* **2022**, *99* (1), 185–194.

(71) Bloom, B. S. Mastery Learning. In *Mastery Learning: Theory and Practice*; Block, J. H., Ed.; Holt, Rinehart, and Winston: New York, 1971; pp 47–63.

(72) Schinske, J.; Tanner, K. Teaching More by Grading Less (or Differently). *CBE Life Sci. Educ.* **2014**, *13* (2), 159–166.

(73) Bloom, B. S. Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment* **1968**, *1* (2).

(74) Guskey, T. R. Making the Grade: What Benefits Students? *Educ. Leadersh.* **1994**, *52* (2), 14.

(75) Aviles, C. B. Grading with Norm-Referenced or Criterion-Referenced Measurements: To Curve or Not to Curve, That Is the Question. *Soc. Work Educ.* **2001**, *20* (5), 603–608.

(76) Prescott, S. G. Will Instructors Save Time Using a Specifications Grading System? *J. Microbiol. Biol. Educ.* **2015**, *16* (2), 298–298.

(77) Shields, K.; Denlinger, K.; Webb, M. Not Missing the Point (s): Applying Specifications Grading to Credit-Bearing Information Literacy Classes. In *The grounded instruction librarian : participating in the scholarship of teaching and learning*; Mallon, M. N., Hays, L., Bradley, C., Huisman, R., Belanger, J., Eds.; Association of College and Research Libraries: Chicago, IL, 2019.

(78) Frey, B. S. How Intrinsic Motivation Is Crowded out and in. *Ration. Soc.* **1994**, *6* (3), 334–352.

(79) Ryan, R. M.; Deci, E. L. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp. Educ. Psychol.* **2000**, *25* (1), 54–67.

(80) Deci, E. L.; Koestner, R.; Ryan, R. M. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Rev. Educ. Res.* **2001**, *71* (1), 1–27.

(81) Pulfrey, C.; Darnon, C.; Butera, F. Autonomy and Task Performance: Explaining the Impact of Grades on Intrinsic Motivation. *J. Educ. Psychol.* **2013**, *105* (1), 39–57.

(82) Cain, J.; Medina, M.; Romanelli, F.; Persky, A. Deficiencies of Traditional Grading Systems and Recommendations for the Future. *Am. J. Pharm. Educ.* **2021**, 8850.

(83) Spidell, C.; Thelin, W. H. Not Ready to Let Go: A Study of Resistance to Grading Contracts. *Composition Studies* **2006**, *34* (1), 35–68.

(84) Peters, R.; Kruse, J.; Buckmiller, T.; Townsley, M. “It’s Just Not Fair!” Making Sense of Secondary Students’ Resistance to a Standards-Based Grading. *American Secondary Education; Bowling Green* **2017**, *45* (3), 9–28.

(85) Covington, M. V. Goal Theory, Motivation, and School Achievement: An Integrative Review. *Annu. Rev. Psychol.* **2000**, *51*, 171–200.

(86) Crocker, J.; Luhtanen, R. K. Level of Self-Esteem and Contingencies of Self-Worth: Unique Effects on Academic, Social, and Financial Problems in College Students. *Pers. Soc. Psychol. Bull.* **2003**, *29* (6), 701–712.

(87) Adams, J. B. What Makes the Grade? Faculty and Student Perceptions. *Teach. Psychol.* **2005**, *32* (1), 21–24.

(88) Zinn, T. E.; Magnotti, J. F.; Marchuk, K.; Schultz, B. S.; Luther, A.; Varfolomeeva, V. Does Effort Still Count?: More on What Makes the Grade. *Teach. Psychol.* **2011**, *38* (1), 10–15.

(89) Greenberger, E.; Lessard, J.; Chen, C.; Farruggia, S. P. Self-Entitled College Students: Contributions of Personality, Parenting, and Motivational Factors. *J. Youth Adolesc.* **2008**, *37* (10), 1193–1204.

(90) Tippin, G. K.; Lafreniere, K. D.; Page, S. Student Perception of Academic Grading: Personality, Academic Orientation, and Effort. *Active Learning in Higher Education* **2012**, *13* (1), 51–61.

(91) Seymour, E.; Melton, G.; Wiese, D. J.; Pederson-Gallegos, L. *Partners in Innovation: Teaching Assistants in College Science Courses*; Rowman and Littlefield: Boulder, CO, 2005.

(92) Yerushalmi, E.; Sayer, R.; Marshman, E.; Henderson, C.; Singh, C. Physics Graduate Teaching Assistants’ Beliefs about a Grading Rubric: Lessons Learned. *arXiv [physics.ed-ph]*, 2017.

(93) Kurdziel, J. P.; Turner, J. A.; Luft, J. A.; Roehrig, G. H. Graduate Teaching Assistants and Inquiry-Based Instruction: Implications for Graduate Teaching Assistant Training. *J. Chem. Educ.* **2003**, *80* (10), 1206.

(94) Sandi-Urena, S.; Gatlin, T. Factors Contributing to the Development of Graduate Teaching Assistant Self-Image. *J. Chem. Educ.* **2013**, *90* (10), 1303–1309.

(95) Mohamed, A.-R. Effects of Active Learning Variants on Student Performance and Learning Perceptions. *International Journal for the Scholarship of Teaching and Learning* **2008**, *2* (2), 11.

(96) Shaw, T. J.; Yang, S.; Nash, T. R.; Pigg, R. M.; Grim, J. M. Knowing Is Half the Battle: Assessments of Both Student Perception and Performance Are Necessary to Successfully Evaluate Curricular Transformation. *PLoS One* **2019**, *14* (1), e0210030.