Implementation of specifications grading in an upper-division chemical biology lecture course

Running title: Specifications grading chemical biology

Jessica I. Kelz[1], Jose L. Uribe[1], Mahtab F. Rasekh[1], Gemma R. Takahashi[2], Wyeth S. Gibson[1], Renée D. Link[1], Kate J. McKnelly[3], and Rachel W. Martin[*1,2]

[1]Department of Chemistry, University of California, Irvine, CA 92697-2025, USA

[2]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697-3900, USA

[3]Department of Chemistry, Emory College of Arts and Sciences, Atlanta, GA 30322, USA

15 March, 2023

[*] Correspondence should be addressed to: rwmartin@uci.edu.

Abstract

Specifications grading is a student-centered assessment method that enables flexibility and opportunities for revision. Here we describe the first known full implementation of specifications grading in an upper-division chemical biology course. Due to the rapid development of relevant knowledge in this discipline, the overarching goal of this class is to prepare students to interpret and communicate about current research. In the past, a conventional points-based assessment method made it challenging to ensure satisfactory standards for student work were consistently met, particularly for the comprehensive written assignments. Specifications grading was chosen because the core tenet requires students to demonstrate minimum learning objectives to achieve a passing grade and complete more content of increased cognitive complexity to achieve higher grades. This strict adherence to determining grades based on demonstrated skills is balanced by opportunities for revision or flexibility in assignment deadlines. These options are made manageable for the instructors through the use of a token economy with a limited number of tokens that students can choose to use when needed. Over the duration of the course a validated survey on self-efficacy showed slight positive trends, student comprehension and demonstrated skills qualitatively improved, and final grade distributions were not negatively affected. Instructors noticed that discussions with students were more focused on course concepts and feedback rather than grades, while overall grading time was reduced. Responses to university-administered student feedback surveys revealed some self-reported reduction in anxiety as well as increased confidence in managing time and course material. Recommendations are provided on how to continue to improve the overall teaching and learning experience for both instructors and students.

INTRODUCTION

Introduction to Chemical Biology is an upper-division course taken by third- and fourth-year undergraduates in the chemistry major at the University of California, Irvine (UCI). It is required for both the chemistry major and to meet the biochemistry requirement for the American Chemical Society (ACS) degree certification[1]. Although the student demographic primarily comprises chemistry majors, the course is also open to students from the School of Biological Sciences as an elective; typical enrollment is around 100-120 students. The course covers the fundamentals of Chemical Biology, specifically the application of chemical techniques and mechanisms to explain biological phenomena at the scale of atoms and bonds. Topics include structures and reactivity, chemical mechanisms of enzyme catalysis, chemistry of signaling, biosynthesis, and metabolic pathways. The lectures provide background information and context required to connect fundamental principles from chemistry with key concepts governing living organisms. In practice, most of the material covered relates to the Central Dogma of Molecular Biology[2], following the flow of information from DNA to RNA to protein. The logic and interpretation of experiments are heavily emphasized in this course; "How do we know?" is at least as important as "What happens?".

Chemical biology has emerged as a recognized subdiscipline within the last several decades and bridges the gap between the molecular detail of chemistry and complex systems of biology. Despite being integral to several areas of transformative research, core competencies such as those outlined for other subdisciplines by the American Chemical Society Committee on Professional Training (ACS CPT) guidelines or seminal texts on undergraduate biology education[3] have not similarly been established for chemical biology[4]. This may be in part because the subject matter is evolving at a very rapid pace[5], making it challenging to develop an integrated curriculum suitable for multiple majors that is appreciable by students and achievable by instructors[6]. For example, the textbook[7] utilized for this course is less than a decade old at present (a short timescale for many STEM subjects), however since the textbook was published in 2013 the genome editing method CRISPR-Cas9[8,9] was developed and subsequently awarded a Nobel Prize, single-molecule benchtop nucleic acid sequencing[10,11] has become commercially available at a price point allowing mass use and mRNA vaccines[12] have been developed for commercial use. This flood of new information is potentially made even more problematic by the "tyranny of the textbook,"[13] as these are often the default learning tool for undergraduate education.

Undergraduate education in such an interdisciplinary subject would benefit greatly from activities or assignments that require students to apply their knowledge to real-world research and mimic responsibilities in future careers. One such activity for upper-division students is the use of case studies that develop critical skills necessary to read literature, justify methods, analyze data, critique findings, and propose hypotheses[4]. Assignments based on peer-reviewed literature need to be well planned so as not to be too complicated or time consuming and are therefore often underutilized in the classroom despite being essential to future education and careers. Not only does addressing this issue this have the potential to ameliorate employer dissatisfaction with recently graduated science major communication skills[14] but it also serves as a means to keep the course material up to date with relevant advances in the field.

The goal of Chem 128 in its most recent iterations (2019-2022) was therefore focused on providing students with a working foundation in chemical biology concepts, techniques, and applications, particularly filtered through the lens of reading the current literature. Central to this objective is the ability to effectively interpret, analyze, and critique scientific papers in writing.

Students are assigned approximately one paper per week from relevant journals, and submit two mini-review assignments during the academic term in which they critique a paper and discuss relevant background literature. The course was taught from 2019-2021 using a traditional points-based assessment system in which the two writing assignments accounted for a total of 20% of the students' final grade. Many students had no prior experience with scientific writing or reading current literature, generating stress for the students and frustration for the instructor. The majority of review papers submitted by students did not meet the expected standards and left the instructor with the unsatisfying choice to either grade the assignment accordingly, which would lower students' grades and be unintentionally discouraging, or give artificially high grades even though the standards were not met. Neither option felt appropriate for the most comprehensive assessments of the course objectives or supportive of student learning. This disconnect motivated the implementation of a simultaneously more rigorous and flexible grading policy.

Specifications grading is a student-centered assessment method focused on demonstration of learning objectives[15]. It has been successfully used in general chemistry lecture[16,17], organic chemistry lecture[18–20], organic chemistry laboratory[21], biochemistry laboratory[22], cell biology lecture[23] and various other STEM courses[24]. Inspired by these efforts, we developed a version of this system for the winter 2022 offering of Chem 128 at UCI. Here we present, to the best of our knowledge, the first implementation of specifications grading in an upper-division chemical biology lecture. Further, we provide a reflective analysis of potential benefits and areas for improvement to future implementations based on student and instructor perceptions and offer considerations for future education research.

SCIENTIFIC AND PEDAGOGICAL BACKGROUND

Proficiency in quantitative analysis is often strongly prioritized in STEM education. However, numerical assessments can be satisfactorily completed without a rigorous conceptual understanding of the material, whereas vague or out of context responses to open-ended questions or essays highlight knowledge deficits [25]. Further, memorization of equations or stand-alone facts does not support the broader goals of science education, which are enabling graduates to apply their fundamental knowledge to make predictions, explain observable outcomes of an experiment, and assess new situations. To the greatest extent possible, information learned should be demonstrated through assessments that mimic real-world use in order to extend the utility of students' knowledge and skills beyond the classroom to independent scholarship[26].

Analytical writing has been demonstrated to enhance conceptual learning, especially when used in tandem with other assignments, to engage the students with material across the cognitive spectrum[27]. Due to the nuanced understanding needed to achieve effective written communication in STEM and its importance to most career paths after graduation, students would likely benefit from pedagogical efforts to incorporate more frequent development of this critical skill[28]. Consistent practice and feedback is most advantageous[22], however written assignments tend to be among the most time-consuming types of assessments to complete and to grade, resulting in less favor amongst both students and instructors. For students, the reasons scientific writing poses a challenge are numerous and multi-faceted. Writing experience gained through other courses such as humanities does not necessarily transfer well due to the distinct organization, specialized terminology, and different audience of lab reports and critiques of peer-reviewed work[29]. More generally, students also tend to have difficulty connecting seemingly disparate knowledge[30], which is then further complicated by simultaneously processing and incorporating new course-specific knowledge, as this is among the highest-level cognitive skills[31].

Simply incorporating more written assessments alone may still not yield the desired results without improved instruction. In order for students to learn content or writing, practice will ideally include elements such as: providing a rationale for the design of an investigation, making sense of data, crafting an argument, and refining a text in light of a critique[32]. Success in these abstract and high-order cognitive tasks is made more challenging by students' complicated relationship with feedback[33–35]. On one hand, students are eager to receive feedback and it is an essential tool for learning. Effective feedback is specific, understandable, and helpful for completing a future task such that a student is willing and able to use it[36]. On the other hand, feedback can also be unintentionally problematic if it is not presented well. Poor-quality feedback is not useful due to being too authoritative, generic, confusing, or if it is unclear how to implement it in future assigments[37]. Although the aforementioned may seem obvious, there are subtleties to successful execution. After receiving a grade, a student may have little motivation to actively engage with the feedback if assignments are viewed as modular[38] or stand-alone products, even if a similar task is assigned later in the course. This lack of incentive is further reinforced if the grade for the assignment has already been determined because students can no longer directly benefit from revision efforts[39]. This contradiction of intent on both sides can be mitigated if the student and instructor use the feedback to create a dialogue such that students are able to incorporate it into their own process of learning[36]. It has been shown that when provided with the opportunity to perform iterative, reflective refinement, student views on feedback improve due to increased literacy and appreciation for the rationale[40]. Proactive recipience, or active engagement in the feedback process[34] is one of the most important factors that increase overall performance[33].

Developing a more flexible and interactive mode of assigning grades also has compelling implications for student learning and inclusion. Traditional grading provides a static picture that is often misconstrued as aptitude, therefore minimizing the opportunity for feedback that could be beneficial to development of creative problem solving. This generally tends to increased anxiety and lower interest in learning, especially among students from minority demographics[41]. Norm-referenced grading was developed because it was believed to be less subjective[42] and is often accepted as a meaningful way to communicate between institutions[43]. However, these "standard" curves can be deceiving because they may represent a comparison of student work relative to each other[44] rather than actually conveying meaningful information about individual student understanding or retention of knowledge[45]. In fact, it has been shown that competitive environments in which students feel the need to outperform peers leads to less retention[46]. Academic performance may become motivated based on extrinsic validation more than intrinsic curiosity, which can impact self-esteem[47] and how students perceive the educational experience in relation to themselves[48]. This does a disservice to students as individuals by denying them effective opportunities to learn through reflection[49,50] as they work toward the ultimate goal of becoming self-regulated learners[51], as well as to the broader scientific community if we are complicit in accepting the loss of talented underrepresented students[52,53], for what at best amounts to tradition given the problems and misconceptions that have been identified. This is particularly important in the wake of the COVID-19 pandemic which disproportionally negatively affected students from minoritized groups[54]. The impact of the pandemic on student well-being will be unique to each individual in terms of its scope and duration[55], however it can potentially be mitigated by efforts in the classroom to improve self-efficacy, a component of well-being that has been correlated to performance. Negative trends in interpersonal communication, problem solving and grades have been reported in a recent study about the return to in-person teaching an institutes of higher education, with a proposed solution being to modify course content and delivery[56].

Specifications grading has the potential to provide several notable benefits for both instructors and students[15]. A specifications grading system was utilized in a "Writing for Chemists" course developed at UCI with the goal of providing students frequent opportunities to engage with feedback and submit revisions[28]. This assessment method differs from the traditional points-based grading system in that students are required to demonstrate achievement of learning objectives at a satisfactory level or no credit is earned. To offset the higher stakes of removing partial credit, a key feature of this method is that instructors must provide very clear, detailed specifications for what is considered satisfactory. For instructors this can result in less time spent grading, and for students this shifts the focus from negotiating partial credit to improving understanding of course concepts in order to adequately demonstrate a learning objective[57]. Also, one of the core tenets of specifications grading is the use of tokens to return a sense of ownership over the learning experience to the students. Tokens provide opportunities for flexibility in submission deadlines and the opportunity to incorporate instructor feedback in the resubmission of revised course assessments while also maintaining a sustainable workload for instructors. To earn higher course grades students must demonstrate a mastery of more advanced or complex skills and content applied to more assignments. Requiring revisions instead of awarding partial credit, motivates students to actively understand why their previous work did not meet learning objectives which supports learning [49,50]. Students will not necessarily achieve all the possible learning outcomes, but their course grade will indicate which outcomes they have and have not achieved. Overall, this method enables instructors to adequately uphold high standards while shifting agency for the overall grade to the student[58–61] by enabling them to revisit challenging concepts or skills in a productive way.

The major goals of the specifications grading redesign of Chem 128 was to promote improvement of the writing assignment submissions such that students could adequately demonstrate application of knowledge to new situations and engagement in scientific argumentation[32], and student self-efficacy through their perceived ability to succeed in the course and confidence to effectively communicate about course concepts. These are both essential skills to advance research literacy and future career success. As we were unable to directly compare other results to previous versions of the course due to the COVID-19 pandemic, these metrics serve as a means of evaluating the effectiveness of this stand-alone implementation toward these goals.

MATERIALS AND METHODS

Course design

Specifications grading can be hybridized with points-based assessments in a partial implementation[17], however we elected to utilize a full-specifications grading option (no points component) in the most recent iteration of the course in order to simplify the assessment policy and to try to create maximum buy-in from the students. This required establishment of rules for using tokens, updates to assignment rubrics to reflect mastery criteria for meeting learning objectives, and creation of an overall grade tracker based on demonstrated proficiency across the various course assessments. The course had several formal assessments over a range of cognitive levels designed to evaluate fundamental understanding of the application of chemical techniques and mechanisms to explain biological phenomena at the scale of atoms and bonds. In previous course iterations these included: discussion section worksheets, problem sets, quizzes, midterm, final, and writing assignments. Minor changes to the grading schema included replacing the two exams with four quizzes because it is our interpretation that high-stakes, summative exams are philosophically contradictory to the intent of specifications grading[15] and eliminating one of the five problem sets due to time constraints. Worksheets, problem sets and quizzes were designed as assessments of fundamental knowledge and skills. The writing assignments were designed as mini reviews of the protein and nucleic acid literature, requiring students to combine concepts learned in the course in order to critically analyze methods, results, and proposed future work.

Token Policy

In this course, students earned all tokens by completing small, course-related activities. Up to seven tokens could be earned over the duration of the quarter broken down as follows: pre-course self-efficacy survey (2), syllabus assignment (1), chemical biology meme (1), attending a relevant department seminar (1) and post-course self-efficacy survey (2); shown as activity and number of tokens earned respectively. The pre-course self-efficacy survey was due by the end of the first week of the class and the post-course survey was due by week eight of the ten-week quarter to provide time to use the earned tokens. Mandatory participation in research-related surveys is prohibited in the classroom so alternative assignments such as reading a chemistry education research publication and writing a brief (2-3 sentences) summary were also made available to students who chose not to participate in the surveys.

The Token Trade-In List provided to students through a page in the course learning management system (LMS) at the beginning of the quarter is provided in the Supplemental

Material. This document detailed specific guidelines on how tokens could be used, which included: resubmission of research paper(s) (first paper 2 maximum, second paper 1 maximum), resubmission of a problem set (2 maximum), revision to one quiz question (1 per quiz, maximum 4), opt to take final to replace quiz score (1 per quiz, maximum 4), not attend a discussion section (1 maximum), and late assignment submission (3 maximum per assignment, 1 token per 24 hour period, 72 hour maximum extension). Maxima that could be applied to any given assessment, a time limit of one week to complete revisions after each assignment, and a deadline to use tokens by week nine of the quarter (except for the final exam) were established as a means to mitigate student and instructor workload. Each problem set and quiz resubmission also required a student reflection on the changes made to correct mistakes or incorporate feedback. Reflections were not required for resubmission of the writing assignments.

The two teaching assistants (TAs) assigned to the course maintained a tracker of tokens earned and used for each student. Individual assignments were marked as either complete or incomplete. TAs then utilized a single, editable "Token" assignment in the LMS, the score for which would increase when tokens were earned and decrease when used in order to monitor the number of tokens each student had available. Students were required to email TAs directly with the specific need (i.e. 24-hour late submission) to request use of tokens. An external inventory was accounted for in an Excel spreadsheet accessible to both TAs which contained how students earned tokens and how they used them.

Rubrics

The writing assignment rubric was adapted from grading criteria used from a writing course taught by K.J.M. at Emory University and previous iterations of the chemical biology course. Updates to and expansion of the rubric made feedback both more general, as it did not require the instructor to provide as many individual comments, and more detailed because each criterion was written to be more specific and clear. Rubric criteria encompassed skills previously observed to be problematic in student scientific writing: scientific vocabulary, concision in writing, formatting and organization, flow, conventions of scientific writing[62], proper use of literature citations, presentation of data, and avoiding plagiarism[14]. Eleven of the twenty-four criteria were designated as “core,” shaded in green in Table 1, and were required to be met along with a cumulative total of 17 for “low pass” and 21 for “high pass” assessment. If the minimum requirements were not met, the assignment was evaluated as “needs revision”. In line with the specifications grading method, criteria beyond those designated “core” were higher-order cognitive tasks such as justification of methods. If minimum criteria to achieve a passing grade were not met, the assignment was marked as “needs revision” and students were allowed the opportunity to apply a token to resubmit. Students who achieved a “low pass” were also permitted to resubmit to attempt to achieve a “high pass”.

Grade Criteria

Ultimately grade criteria are at the discretion of the instructor, which maintains academic freedom in applying this method. However, the general expectation in specifications grading is that students will need to demonstrate mastery of skills or concepts with higher cognitive demand and / or complete more work in order to earn higher final letter grades. We used Bloom’s taxonomy[63,64] to establish baseline skills for grade demarcations. Each question on a problem set or quiz was assigned a letter grade for the purpose of establishing performance thresholds on assignments. “C”-level questions were based on knowledge and understanding, requiring students to: define, summarize, identify, and perform simple calculations. “B”-level questions were based on application and analysis, requiring students to: make connections among different topics, apply principles to a new problem, draw structures, propose mechanisms, or deduce the correct equations to use. “A”-level questions were based on evaluation and creating, requiring students to: explain how methods were used, justify methods and controls by assessing their impact on the results, generate hypotheses and describe an experimental design to test them, or make predictions. These general descriptions were made available to the students, however the letter grade associated with each question was not released until afterwards in order to promote maximum participation in the exercises. Minima for low pass and high pass scores were consistently applied to all assignments and quizzes. To earn a low pass students were required to either satisfactorily complete all of the “C”-level questions or all but one of the “C”-level questions and at least one other question. To earn a high pass students were required to demonstrate at least all but one of the “C”-level questions and achieve at least 80% satisfactory completion of the assignment, which would necessitate demonstrated skills at both the “B” and “A”-level. If the criteria for low pass were not met then the assignment or quiz would be returned as “needs revision” and the student would be allowed to use to token to preform revisions and improve the score. The highest score achieved after allowed resubmissions was recorded.

The overall grade determination matrix for the course is presented in Table 2. Students earned the highest grade for which they met all of the minimum requirements. In order to achieve a “D”, students were required to earn a “low pass” on all assessments and complete six discussion section worksheets. Plus and minus grades are used at UCI, so additional distinctions were made from the base grade requirements. For plus grades students needed to complete at least one additional discussion section worksheet and earn a “high pass” on a research paper when “low pass” was required. For minus grades, students were permitted completion of one fewer discussion section worksheet and earning a “low pass” when “high pass” was required on a research paper.

Self-Efficacy Survey

The fourteen-question self-efficacy survey used for this course, provided in the Supplemental Material, was modified from a validated survey to probe student confidence in learning biology, especially as non-majors[65]. There were three assessment factors addressed by the questions: methods of chemical biology (question 1), generalization to other chemical biology/science courses and analyzing data (questions 2-7), and application of chemical biology concepts and skills (questions 8-13). The survey questions were adapted very minimally to make the wording applicable to this course. Table 3 documents the changes in wording from the original survey questions (bold) to the survey used for this course (bold, italic). Question 14 was the only question we added that was not adapted from the original survey but was deemed pertinent to assessing the goals of the course. The full survey is provided in the Supplemental Materials for further reference. The survey was made available through the UCI’s instance of Qualtrics, a cloud-based platform for distributing web-based surveys. Participation was completely voluntary (an alternative assignment was provided for students who chose not to participate) and results were analyzed *en masse* to maintain anonymity.

RESULTS AND DISCUSSION

A total of 107 students enrolled and 99 students completed the winter 2022 iteration of the course described here. We judged the use of specifications grading to be an overall success, as there were no concerning differences in overall grade distribution, the mean results of the student self-efficacy survey improved slightly, and there were substantial improvements observed on several rubric metrics between the initial submission of writing assignment 1 and writing assignment 2. This is particularly significant because it was many students' first exposure to this grading method which can initially cause anxiety[21,66] and it was the first implementation for this course which can be challenging for a variety of reasons[67]. We are encouraged by these results that other educators in biophysics may be able to adapt this framework for their own classrooms.

Token Economy

The token system should ideally be aligned to support demonstrated mastery of course objectives without allowing students to generate an unmanageable workload for themselves or the instructors[15,68]. Providing too few tokens causes students to hoard them, preventing them from revising their work, whereas providing too many allows students to mismanage their workload by pushing everything to the end of the class, which is a suboptimal learning experience as well as producing an unrealistic amount of grading for the instructors at the end of the course. We designed our token economy similar to the system implemented in the "Writing for Chemists" course[28].

Tracking tokens not only served as a means of accounting but also allowed for analysis of the overall way students used their tokens. Out of seven total available, the average number of tokens earned and used was six and four respectively. Thirty-five out of ninety-nine students used fewer than half of the available tokens and only four used all seven. As shown in Table 4, the highest percentage of tokens were used on writing assignment 1 (124, 32.9%), quiz revisions (103, 27.3%), and writing assignment 2 (66, 17.5%). While exact replication of this policy is not the only means to achieve these results, as administered the token system employed adequately supported the goals of the course as it was not detrimental to student performance or instructor workload.

Writing Assignments

Using specifications rubrics for the writing assignments in particular enables students to learn from their mistakes on this challenging and novel (for them) task in a low-stakes context. The nucleic acid mini-review paper was assigned in week 4 of the ten-week quarter and students were allowed to use tokens to resubmit up to two times. The protein mini-review paper was assigned in week 8 and students were allowed to use tokens to resubmit once due to time constraints at the end of the quarter. Two students did not submit either assignment despite having access to tokens that could have enabled a late submission. A detailed breakdown of the criteria marked as “needs revision” for the initial submission and any resubmissions for each writing assignment is provided in Table 5, where criteria shaded in green and marked with an asterisk are core. Bolded red / negative values indicate more than 25% of the class did not adequately demonstrate the rubric line item. Bolded green / positive values indicate criteria with the largest amount of improvement (less frequently marked as “needs revision”) between writing assignment 1 (WA1) and writing assignment 2 (WA2). Five overall criteria comprised of four core (citations format and placement, discussion, controls, conclusions) and one other (clarity) were marked as “needs revision” for 25% or more of the class on initial submissions for both writing assignments. Criteria that showed the most improvement from the initial submission of WA1 to the initial submission of WA2 were discussion, controls and technical writing which improved by 28%, 32% and 75% respectively, indicating that learning improved between the two assignments. In total fourteen (eleven not previously mentioned) of the twenty-four criteria yielded a decrease in the frequency of “needs revision” evaluations between the initial submissions of both assignments. Criteria where students did not improve between the initial submission of writing assignments were relatively anomalous, impacting less than 10% of the students, however this information could indicate areas to be emphasized with additional practice or discussion in future iterations of the course.

For both writing assignments, most students received an overall evaluation of “needs revision” on the first submission, but achieved “high pass” by the final submission, as shown in Table 5. Slightly more students received a final grade of “low pass” on the second paper, likely due to only having one submission attempt and possibly other competing time requirements at the end of the quarter. The reason we do not assess this to represent declining performance is because roughly 20% of students improved the initial submission grade from writing assignment one to two, with “needs revision” dropping from 87 to 65 respectively. Students not only applied feedback to make corrections to each individual assignment, but these results indicate that feedback from WA1 was also utilized to improve the initial submission of WA2. We interpret this finding to demonstrate that students learned new skills and knowledge throughout the revision process. Almost all students were able to achieve “high pass” on both writing assignments, and while not directly comparable to previous iterations of the course, student performance was qualitatively noted to be much more consistent and improved overall.

Grade Distributions

This course was taught by the same instructor for four consecutive years beginning in winter quarter of 2019. In 2019, students’ final letter grades were determined by the total points accumulated over the duration of the course from the following assessments: quizzes and discussion problems (10%), problem sets (15%) writing assignments (20%), midterm (25%), and final exam (30%). The late policy for points-based grades permitted assignments to be accepted up to one hour late with no penalties and a 10% reduction in score for assignments received each 24-hour period beyond the original deadline. While using points-based assessments, students were not permitted to revise or resubmit work. Specifications grading was utilized in 2022 with the grade criteria and token policies previously described.

Final grade distributions for the 2019 and 2022 courses are shown in SM Figure 1. Winter 2020 and 2021 grades were omitted from the comparison because these iterations were substantially altered to accommodate remote instruction due to the COVID-19 pandemic. The 2019 points-based grade distribution was characteristically Gaussian with a mode grade of B+ (typical for an upper-division course taken primarily by majors) for a class size of 108 students. In this implementation of specifications grading, significantly more students earned A+ and A final grades, yielding a unimodal distribution across the 99 students. The net workload and expectations for the course predominantly remained unchanged. Therefore, the grade shift is representative of more students demonstrating mastery of the learning objectives, in part due to opportunities for revision. As an example of this, make-up quizzes were written to be conceptually similar but with unique questions such that answers could not be memorized and learning must be demonstrated. The general shift to higher grades is consistent with some other implementations of specifications grading in undergraduate STEM education[16,21,69] We hypothesize that this may be in part because a student that would typically earn a "B" in a traditional points-based system is presented with the tools and awareness to achieve an "A"[16,45,70]. The grade distributions are not directly comparable to each other in terms of changes in student learning due to adjustments in the course structure and the unknowable impact of the COVID-19 pandemic. However, we have included the grades to provide a baseline for evaluating whether we provided enough opportunities for rework and in order to demonstrate this implementation did not lower students' grades on average despite the more rigorous standards.

Survey Results

We surveyed students at the beginning and end of the course to test whether student perceptions about their ability to succeed in this or related courses improved after exposure to the more self-directed learning approach offered in specifications grading, or alternatively, if it declined due to receiving detailed, critical feedback. As determined by the token tracker, one student did not complete either survey, twenty-two students (some of whom dropped the course) only completed the first survey, and two students only completed the second survey. Sixteen students submitted two entries for one or both of the surveys, possibly by mistake, therefore we elected to include only the first response in the analysis. This was determined based on IP address alone as names were used only for awarding token credit and were removed from the survey results prior to analysis. In total 77 sets of surveys (~78%) were used in this investigation.

Students responded to the 14 questions with a Likert-scale ranging from (1) NOT AT ALL confident to (5) TOTALLY confident[65]. Results of the pre-course (week 1) and post-course (week 8) surveys were paired for each student. The mean result was determined for the question(s) corresponding to each assessment factor for each set of surveys[65,71]. Student response means for each of the three original factors as well as the question we added were assessed for statistically significant changes. We performed both paired t-tests and Wilcoxon signed-rank tests in R statistical software[72,73] to determine whether results were significant. The results of the paired t-tests for each factor are provided in Table 7 and distributions of the initial and final factor averages are presented in Figure 1. Both tests qualitatively validated that confidence in all factors increased, indicating that student self-efficacy improved over the duration of the course. The results of this survey demonstrated that specifications grading qualitatively improved student perceptions on self-efficacy to succeed in the course and communicate about related topics, especially in areas of particular focus related to the goals of the class. Extensive prior research has focused on the influence of mindset on academic performance. Our results corroborate this relationship and further suggest that academic performance influences students' mindsets.[74]

Limitations of this study are mostly due to its being the first implementation of specifications grading in this course. For instance, we did not include a control group, in part because this was the first implementation of specifications grading in the course and only one section of the class was offered during that quarter. In the future it would be beneficial to perform the survey in the same manner with a version of the class with the same assessments and rubrics but taught using a traditional points-based system. We did not receive responses for both surveys from every student enrolled throughout the course, so it is possible that students who were already biased toward feeling confident answered. Further, the questions are a qualitative self-reflection which may be impacted by many factors outside of administration of this course.

Student Perceptions

University-administered teaching evaluations were completed by 29/99 students at the end of the quarter. The free-response questions used the standard wording for teaching evaluations at UCI and therefore did not ask about specifications grading in particular. These questions are:

1. Which aspects of this class did you feel were intellectually or creatively stimulating?
2. Which aspects of this class did you feel contributed most to your learning?
3. Which aspects of this class could be improved to enhance your learning?

Here we summarize the responses to these questions that related to specifications grading aspects of the course. Comments on other course features, such as the specific topics covered, the lectures, or the discussion sections, are not included. Students' comments on specifications grading in this course were mostly positive, and many of the negative comments focused on organizational issues related to this being the first time the grading scheme was implemented in this course.

Students liked that the course was organized around four quizzes rather than a midterm and a final. Some found it easier to stay engaged and monitor their progress with more frequent assessments. Reduced anxiety due to the lower stakes of each quiz was also mentioned. Although more frequent, low-stakes assessments are not unique to any one grading method, they are essentially required by specifications grading in order to adequately allow opportunities for rework. Students appreciated the increased transparency afforded by specifications grading, since they knew from the beginning how their grades would be determined. They also found that specifications grading made it easier to understand what to prioritize, which is important in a class where a large amount of complex material is covered. Some students appreciated completing revisions, which allowed the opportunity to learn from mistakes, and the token economy, which enabled management of revision attempts. Of the 28 respondents, 72% answered that the instructor provided opportunities to better understand material (36% strongly agree, 36% agree). These results are consistent with expected benefits of specifications for the student learning experience[15].

Students also provided suggestions for improvement, many of which focused on the materials being new and not previously tested. As an example of relatively common feedback[21,75], some students found the rubrics confusing and thought the grading scheme could be explained better. We plan to improve these materials for future use based on the students' comments. Some other requests are more difficult to implement or are inconsistent with course goals. For example, one student mentioned wanting to know which questions are "A," "B," or

"C" before the assignment is turned in. We made the deliberate choice not to reveal the question classifications until after the assignment is turned in because we wanted students to make a good-faith effort on all problems rather than only attempting the "C" or "C and B" problems. Some students wanted more time to revise the assignments, and one specifically requested an unlimited window until the end of the quarter. Although we will be more mindful of spreading out the assignments in the future, it is not realistic or desirable to offer unlimited time for revisions, both because of the instructional team's workload, and because allowing assignments to pile up until the end of the quarter rather than revising them in a timely manner does not provide an optimal learning experience for students. Finally, one student expressed dislike for specifications grading because it is more work for the students, particularly those without substantial writing practice. However, they also acknowledged understanding our goals in implementing it and voiced that they felt it made them a better writer which is consistent with student perceptions in other writing classes utilizing specifications grading[28,76] and is consistent with the more general observation of student dissatisfaction with methods they view as unconventional regardless of improved performance[77].

Teaching Assistant and Instructor Perceptions

Here we present qualitative assessments assembled from the teaching assistants and course instructor following completion of the grade submissions. From an instructional standpoint, it was expected that some challenges would arise due to this being the first implementation of specifications grading for the course and this grading scheme being new to many students. After a brief initial period of clarifying the instructions related to grading rubrics and token use, the majority of student interactions at office hours and after class meetings were focused on substantive topics related to learning objectives such as how to identify the controls in an experiment or how to draw a chemical mechanism correctly. From an instructor perspective, the best feature of specifications grading was the shift in focus from points and grades to problem-solving and skills. It was observed that less time was dedicated to discussing grades because the overall course expectations were generally clearer, with a path to achieve a given letter grade and all assignments either satisfactory or returned as "needs revision". This was a welcome contrast from previous versions of the same course, where most discussions were concentrated on negotiating for more partial credit and discussing how many points were lost for particular mistakes without the ability to directly correct them, making feedback frustrating for the students and the instructor. Removing the possibility of partial credit seemed to shift the conversation in a more productive direction, toward mastering the skills needed to succeed at the writing assignments or quizzes. This is not always the case with point-based systems where partial credit can contribute significantly to accumulating enough points to achieve a desired overall grade[15,57,78], or where final grades may ultimately be subject to curves or weighted adjustments in order to achieve a desired distribution. As a positive and perhaps non-intuitive outcome for instructors, grading was much more straightforward and faster even when accounting for time spent grading resubmissions. Open-ended questions were still challenging because a key or rubric cannot fully capture every possible variation of a correct answer or a formatting issue so some discernment is required. However, this would be the case in a points-based system as well, and it may be even more challenging to fairly apply partial credit, whereas if instructors are in doubt in specifications grading it is fully appropriate to mark as "needs revision" and allow informed revision. Adoption of this line of thinking can be challenging even with substantial buy-in, because teaching assistants and instructors have all been indoctrinated almost exclusively to points-based systems. During the course, one teaching assistant was concerned that the binary nature of specifications grading as either a pass or needs revision could be detrimental to student grades. Student communication with TAs and the course instructor was observed to improve, generally noted as more positive, less anxious, more eager to improve, and more focused on course concepts.

Considerations for Future Implementation

Buy-in from teaching assistants is critical to realize the benefits to both students and instructors. In this case, even though both teaching assistants (TAs) understood and supported the goals of specifications grading, they still found it difficult to grade each question in a binary manner after previous experiences with assigning partial credit. This required occasional reminders during our regular instructional team meetings to grade quickly and assign a passing score only when **all** required elements of the correct answer were present. In between these discussions, it was easy for TAs to slip back into the default mode of thinking about partial credit, which is contrary to the course goals and takes up too much of the TAs' time. The latter point is especially critical when dealing with revisions: because each assignment may be graded more than once, the workload becomes unmanageable if grades are not assigned quickly and without considering student effort or trying to rationalize partially correct answers. This was mostly a concern at the beginning of the course and became less of a problem with practice. Overall, the TAs, one of whom had taught the same course before the implementation of specifications grading, reported that the average workload for this course was about the same as for similar courses. The issues with implementation could potentially be mitigated by incorporating a brief training for TAs, especially those not or less familiar with specifications grading, before the course begins.

Based on some core criteria of the writing assignments being consistently rated as "not met" for the majority of students on initial submissions, shown in Table 5, it could be beneficial to break these criteria down and incorporate consistent practice into problem sets. Questions based on reading a piece of literature were included in a few problem sets but it may be beneficial to include them on all problem sets in the future. The questions also could be more clearly related to the core criteria on the writing assignment rubrics, which may then help students make the connection between the problem sets and the writing assignments. One other idea to support improvement in this area was to provide students with examples of acceptable assignments, however the instructor determined that this was not aligned with the learning objectives. The students are presented with several exemplars of well-written, brief review papers (e.g. Nature "News and Views,") throughout the course. However, they are not provided with examples of this particular assignment because the goal is for them to analyze and discuss the assigned papers based on their own understanding rather than simply following a template. Further clarification to rubric line items based on student questions and feedback is likely to continue to be important in any future implementations of specifications grading due to the all-or-nothing credit system.

In this implementation answer keys for problem sets and quizzes were posted immediately after initial grades were released to students and reflections for resubmitted quizzes and problem sets were not required to be in a specific format. In the future, to ensure that the resubmission demonstrates learning and mastery of a learning objective we plan to require students to answer the following prompts in addition to the correct answer for each question to be reassessed: "1. What was incorrect about the first approach or answer? Briefly explain why. 2. What changes did you make to achieve the correct answer? Briefly explain why these changes were necessary. 3. What did you learn that you will apply to problems like this in the future?"

We hope that questions will require students to actively re-engage with the course material and reassess any misunderstandings, promote long-term retention of the material.

It is expected that a handful of outliers may not meet all required criteria as presented in the grade determination matrix. It is not realistic to predict every possible scenario that could lead to this, however it is beneficial to have a strategy to mitigate this as uniformly as possible. In this course, most of the observed grading challenges arose when students did not meet all of the specifications needed to earn a low pass for the second writing assignment after one round of feedback and revision. Ideally, they would have a second opportunity to revise their work and earn a better grade; however, this was not feasible because it was too close to the end of the course. In all four cases where this happened, the students' second drafts showed significant improvement relative to the first, and they were assigned a score of low pass, enabling them to pass the course. One other student turned in a "revised" second writing assignment without having submitted the first draft; this was graded normally and earned a score of high pass. Although improving the rubrics and instructions will likely reduce the number of exceptions that have to be dealt with, it is probably impossible to eliminate them altogether and some flexibility is needed to determine grades in these cases.

The only major drawback of this implementation of specifications grading was the accumulation of grading near the end of the quarter. In particular, two rounds of revisions were allowed for the first writing assignment in order to make sure students were provided with enough feedback on their work and opportunities to correct mistakes. However, the initial submission for the first writing assignment was late enough in the quarter that the second round of revisions coincided with the initial submission of the second writing assignment, causing a bottleneck in grading. This led to excessive work for the instructor during this time, as well as a delay in students' receiving feedback. We believe this problem can be resolved with better scheduling, particularly moving the first writing assignment earlier in the quarter, even though students will not have as much background when they begin to work on it.

CONCLUSION

Due to the rapid pace of changes in the field of chemical biology, an upper-division undergraduate course was redesigned using specifications grading to support research literacy as demonstrated through comprehensive writing assignments. Specifications grading offers a tailorable, student-centered assessment approach that can be beneficial for both students and instructors, especially for high-complexity cognitive tasks that can benefit from iterative feedback. The grading system allowed students to resubmit work, qualitatively improving both their conceptual understanding and their written communication skills. Students overall were receptive to the changes and showed improvements in both self-efficacy and performance in areas aligned with the course learning objectives. Workload for the instructors was comparable to past versions of the course. Although this system requires some buy-in and additional efforts at clarification, it is likely to be beneficial in other interdisciplinary and dynamic areas of study.

IRB STATEMENT

This work, which is classified as exempt (research involving normal education practices in an established educational setting), was carried out in accordance with the standards established by the UC Irvine Institutional Review Board (UCI IRB protocol number 264).

AUTHOR CONTRIBUTIONS

R.W.M. was instructor of record for the course and graded the writing assignments. R.W.M. and J.I.K. co-designed the specifications grading criteria and rubrics, set-up the self-efficacy survey, performed analysis of survey results, token usage, and student feedback, and assembled the

Supplemental Materials. J.LU. and M.F.R. were teaching assistants for the class and were responsible for grading problem sets and quizzes, tracking tokens earned and used, and providing TA perspectives for the manuscript. G.R.T. contributed to statistical analysis of survey results and W.S.G. assisted with development of the violin plots presented in the Supplemental Materials. R.D.L. assisted with the statistical analysis and presentation of the survey results presented in the main text. R.D.L. and K.J.M. provided consultation on implementation of specifications grading and frameworks for course administration before and during the course. J.I.K. and R.W.M. wrote the manuscript.

ACKNOWLEDGMENTS

This research was supported by NSF award DMR-2002837 to R.W.M. and D.J. Tobias and the California Education Learning Lab Grant Project Title: The Teaching Experiment Academy
OPR Issued Grant Number: OPR19178. J.I.K. acknowledges support from the Department of Chemistry and the School of Physical Sciences at UCI. The funders had no role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript. We thank Kailey Baez and Bo Choi from UCI DTEI for practical advice about specifications grading implementation. Most importantly, we gratefully acknowledge the hard work and helpful input of the students in Chem 128, Winter 2022 (UC Irvine).

REFERENCES

1. ACS guidelines and evaluation procedures for bachelor's degree programs, American Chemical Society Office of Professional Training: Washington, D.C., 2015.

2. Crick, F. 1970. Central dogma of molecular biology. Nature 227:561-563.

3. BIO2010: transforming undergraduate education for future research biologists, National Academies Press: Washington, D.C., 2003.

4. Van Dyke, A.R., Gatazka, D.H., and Hanania, M.M. 2018. Innovations in undergraduate chemical biology education. ACS Chem. Biol. 13:26–35.

5. Godwin, H.A., and Davis, B.L. 2005. Teaching undergraduates at the interface of chemistry and biology: challenges and opportunities. Nat. Chem. Biol. 1:176–179
.

6. Begley, T.P. 2005. Chemical biology: an educational challenge for chemistry departments. Nat. Chem. Biol. 1:236–238.

7. Van Vranken, D.L., and Weiss, G.A. Introduction to bioorganic chemistry and chemical biology; Garland Science, 2013.

8. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821.

9. Doudna, J.A., and Charpentier, E. 2014. The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096.

10. Wang, Y., Zhao, Y., Bollas, A., Wang, Y., and Au, K.F. 2021. Nanopore sequencing technology, bioinformatics and applications. Nat. Biotechnol. 39:1348–1365.

11. Wang, Y., Yang, Q., and Wang, Z. 2015. The evolution of nanopore sequencing. Front. Genet. 5:10.3389/fgene.2014.00449.

12. Pardi, N., Hogan, M.J., Porter, F.W., and Weissman, D. 2018. mRNA vaccines — a new era in vaccinology. *Nat. Rev. Drug Discov*. 17:261–279.

13. Allen, D., and Tanner, K. 2007. Putting the horse back in front of the cart: using visions and decisions about high-quality learning experiences to drive course design. CBE—Life Sci. Educ. 6:85–89.

14. Bailey, D.N., and Markowicz, L. 1983. Chemistry and English: a new bond. J. Chem. Ed. 60:467-468.

15. Nilson, L.B., and Stanny, C.J. Specifications grading: restoring rigor, motivating students, and saving faculty time; Stylus Publishing, 2015.

16. Hollinsed, W.C. 2018. Applying innovations in teaching to general chemistry. ACS Symposium Series: Increasing Retention of Under-Represented Students in STEM through Affective and Cognitive Interventions. 145–152.

17. Martin, L.J. 2019. Introducing components of specifications grading to a general chemistry I course. ACS Symposium Series: Enhancing Retention in Introductory Chemistry Courses: Teaching Practices and Assessments.105–119.

18. Ring, J. 2017. ConfChem Conference on Select 2016 BCCE Presentations: Specifications grading in the flipped organic classroom. J. Chem. Educ. 94:2005–2006.

19. Houseknecht, J.B., and Bates, L.K. 2020. Transition to remote instruction using hybrid just-in-time teaching, collaborative learning, and specifications grading for organic chemistry 2. J. Chem. Educ. 97:3230–3234.

20. Ahlberg, L. 2021. Organic chemistry core competencies: helping students engage using specifications. ACS Symposium Series: Engaging Students in Organic Chemistry. 25-36.

21. Howitz, W.J., McKnelly, K.J., and Link, R.D. 2021. Developing and implementing a specifications grading system in an organic chemistry laboratory course. J. Chem. Educ. 98:385–394.

22. Ponce, M.L.S., and Moorhead, G.B.G. 2020. Developing scientific writing skills in upper level biochemistry students through extensive practice and feedback. FASEB J. 34:1–1.

23. Katzman, S.D., Hurst-Kennedy, J., Barrera, A., Talley, J., Javazon, E., Diaz, M., and Anzovino, M.E. 2021. The effect of specifications grading on students' learning and attitudes in an undergraduate-level cell biology course. J. Microbiol. Biol. Educ. 22:e00200-21.

24. Tsoi, M.Y., Anzovino, M.E., Erickson, A.H.L., Forringer, E.R., Henary, E., Lively, A., Morton, M.S., Perell-Gerson, K., Perrine, S., Villanueva, O., Whitney, M., and Woodbridge, C.M. 2019. Variations in implementation of specifications grading in STEM courses. GA J. Sci. 77:Article 10.

25. Nakhleh, M.B. 1994. Chemical education research in the laboratory environment: how can research uncover what students are learning? J. Chem. Educ. 71:201.

26. Larkin, J.H., and Reif, F. 1976. Analysis and teaching of a general skill for studying scientific text. J. Educ. Psychol. 68:431–440.

27. Applebee, A.N. 1984. Writing and reasoning. Rev. Educ. Res. 54:577-596.

28. McKnelly, K.J., Morris, M.A., and Mang, S.A. 2021. Redesigning a "writing for chemists" course using specifications grading. J. Chem. Educ. 98:1201–1207.

29. Erduran, S., and Aleixandre, M. Argumentation in science education: perspectives from classroom-based research; Springer, 2008.

30. Bergmann, L.S., and Zepernick, J. 2007. Disciplinarity and transfer: students' perceptions of learning to write. Writing Program Administration 31:124-149.

31. Hodges, L.C. Teaching undergraduate science: a guide to overcoming obstacles to student learning; Stylus Publishing, 2015.

32. Sampson, V., and Walker, J.P. 2012. Argument-driven inquiry as a way to help undergraduate students write to learn by learning to write in chemistry. Int. J. Sci. Educ. 34:1443–1485.

33. Winstone, N.E., Nash, R.A., Parker, M., and Rowntree, J. 2017. Supporting learners' agentic engagement with feedback: a systematic review and a taxonomy of recipience processes. Educ. Psychol. 52:17–37.

34. Winstone, N.E., Nash, R.A., Rowntree, J., and Parker, M. 2017. 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. Stud. High. Educ. 42:2026–2041.

35. Gibbs, G., and Simpson, C. 2005. Conditions under which assessment supports students' learning. Learn. Teach. in High. Educ. 1:3-31.

36. Price, M., Handley, K., Millar, J., and O'Donovan, B. 2010. Feedback : all that effort, but what is the effect? Assess. Eval. High. Educ. 35:277–289.

37. Jonsson, A. 2013. Facilitating productive use of feedback in higher education. Act. Learn. High. Educ. 14:63–76.

38. Price, M., Handley, K., and Millar, J. 2011. Feedback: focusing attention on engagement. Stud. High. Educ. 36:879–896.

39. Mann, S.J. 2001. Alternative perspectives on the student experience: alienation and engagement. Stud. High. Educ. 26:7–19.

40. Hill, J., and West, H. 2020. Improving the student learning experience through dialogic feed-forward assessment. Assess. Eval. High. Educ. 45:82–97.

41. Schinske, J., and Tanner, K. 2014. Teaching more by grading less (or differently). *CBE—Life Sci. Educ.* 13:159–166.

42. Guskey, T.R. 1994. Making the grade: What benefits students? Educational, School, and Counseling Psychology Faculty Publications (13), University of Kentucky.

43. Schneider, J., and Hutt, E. 2014. Making the grade: a history of the A–F marking scheme. *J. Curric. Stud.* 46:201–224.

44. Brookhart, S.M. Grading; Pearson, 2004.

45. Bloom, B.S. 1968. Learning for mastery. Evaluation Comment, Center for the Study of Evaluation of Instructional Programs, University of California at Los Angeles. 1:1-11.

46. Humphreys, B., Johnson, R., and Johnson, D. 1982. Effects of cooperative, competitive, and individualistic learning on students' achievement in science class. *J Res Sci Teach* 19:351–356.

47. Crocker, J. 2002. The costs of seeking self-esteem. *J. Soc. Issues* 58:597–615.

48. Rogers, C. On becoming a person: A therapist's view of psychotherapy; Constable, 2004.

49. Jonsson, A., and Svingby, G. 2007. The use of scoring rubrics: Reliability, validity and educational consequences. *Educ. Res. Rev.* 2:130–144.

50. Reddy, Y.M., and Andrade, H. 2010. A review of rubric use in higher education. *Assess. Eval. High. Educ.* 35:435–448.

51. Nicol, D.J., and Macfarlane-Dick, D. 2006. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. High. Educ.* 31: 199–218.

52. Seymour, E. and Hewitt, N.M., Talking about leaving: Why undergraduates leave the sciences; Westview Press, 1997.

53. Tobias, S. They're not dumb, they're different: Stalking the second tier; Research Corporation, 1990.

54. Education in a pandemic: The disparate impacts of COVID-19 on America's students, Department of Education Office for Civil Rights, 2021.

55. Plakhotnik, M.S., Volkova, N.V., Jiang, C., Yahiaoui, D., Pheiffer, G., McKay, K., Newman, S., and Reißig-Thust, S. 2021. The perceived impact of COVID-19 on student well-being and the mediating role of the university support: Evidence from France, Germany, Russia, and the UK. *Front. Psychol.* 12: 642689.

56. Becker, T.B., Fenton, J.I., Nikolai, M., Comstock, S.S., Swada, J.G., Weatherspoon, L.J., and Tucker, R.M. 2022. The impact of COVID-19 on student learning during the transition from remote to in-person learning: using mind mapping to identify and address faculty concerns. *Adv. Physiol. Educ.* 46:742–751.

57. Toledo, S., and Dubas, J.M. 2017. A learner-centered grading method focused on reaching proficiency with course learning outcomes. J. Chem. Educ. 94:1043–1050.

58. Docan, T.N. 2006. Positive and negative incentives in the classroom: an analysis of grading systems and student motivation. JoSoTL. 6:21-40.

59. Diegelman-Parente, A. The use of mastery learning with competency-based grading in an organic chemistry course. J. Coll. Sci. Teach. 40:50-58.

60. Voorhees, R.A. 2001. Competency-based learning models: a necessary future. New Dir. Institutional Res. 2001:5–13.

61. Eller, E. 2016. Grading rigor in counselor education: a specifications grading framework Educ. Res. Q. 39:21-42.

62. Wackerly, J.Wm. 2018. Stepwise approach to writing journal-style lab reports in the organic chemistry course sequence. J. Chem. Educ. 95:76–83.

63. Airasian, P.W., Raths, J., Wittrock, M., Mayer, R.E., Pintrich, P., and Cruikshank, K.A. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. Addison Wesley Longman, Inc., 2001.

64. Bloom, B.S. Taxonomy of educational objectives: the classification of educational goals. D. McKay, 1974.

65. Baldwin, J.A., Ebert-May D., and Burns, D.J. 1999. The development of a college biology self-efficacy instrument for nonmajors. Sci. Ed. 83:397-408.

66. Helmke, B. 2019. Specifications grading in an upper-level BME elective course. *ASEE Conferences*, 33278.

67. Streifer, A.C., and Palmer, M.S. 2021. Is specifications grading right for me?: A readiness assessment to help instructors decide. *Coll. Teach*., 1–8.

68. Doll, C., McLaughlin, T.F., and Barretto, A. 2013. The token economy: a recent review and evaluation. Int. J. Basic Appl. Sci. 02:131-149.

69. Carlisle, S. 2020. Simple specifications grading. PRIMUS 30:926–951.

70. Damavandi, M.E., and Shekari Kashani, Z. 2010. Effect of mastery learning method on performance, attitude of the weak students in chemistry. Procedia - Soc. Behav. Sci. 5:1574–1579.

71. Youngblood, J.P., Webb, E.A., Gin, L.E., van Leusen, P., Henry, J.R., VandenBrooks, J.M., and Brownell, S.E. (2022). Anatomical self-efficacy of undergraduate students improves during a fully online biology course with at-home dissections. Adv. Physiol. Educ. *46*, 125–139. 10.1152/advan.00139.2021.

72. R Core Team, R. A language and environment for statistical computing [Online]. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

73. 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. 2019. Welcome to the Tidyverse. J. Open Source Softw. 4:1686.

74. Limeri, L.B., Carter, N.T., Choe, J., Harper, H.G., Martin, H.R., Benton, A., and Dolan, E.L. 2020. Growing a growth mindset: characterizing how and why undergraduate students' mindsets change. Int. J. STEM Educ. 7:35.

75. Fernandez, T., Martin, K., Mangum, R., and Bell-Huff, C. 2020. Whose grade is it anyway?: Transitioning engineering courses to an evidence-based specifications grading system. *ASEE Conferences*, 35512.

76. Mirsky, G. M. 2018. Effectiveness of specifications grading in teaching technical writing to computer science students. J. Comput. Sci. Coll. 34:104-110.

77. Deslauriers, L.; McCarty, L. S.; Miller, K.; Callaghan, K.; Kestin, G. 2019. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc. Natl. Acad. Sci. 116:19251–19257.

78. Blackstone, B., and Oldmixon, E. (2019). Specifications Grading in Political Science. J. Polit. Sci. Educ. 15, 191–205.

SUPPLEMENTAL MATERIAL

The supplemental material contains the token trade-in document provided to students, grade criteria, grade distributions, self-efficacy survey, and specifics of self-efficacy survey statistical analysis.

FIGURES AND TABLES

Figure 1: Results of student self-efficacy survey recorded in week 1 (pre-course) and week 8 (post-course) of the Winter 22 quarter (n=77). Responses range from (1) NOT AT ALL confident to (5) TOTALLY confident. Means are depicted as a gray dot within the boxed interquartile range. All assessment factors had a significant and positive change in the mean response from the initial to final survey.

Table 1: Assessment Criteria for Writing Assignments

Writing Assignment Rubric Criteria	
Page Limit: Is within 1/2 page of the limit and does not exceed the maximum number of pages.	**Spelling:** Words are spelled and used correctly (contains fewer than 2 errors).
Citations: References are cited using the format of any journal. Author names or first author et al. must be included, along with title, journal, issue, page numbers or article number, and year. Papers are cited in the order they are mentioned, and figure captions include citations for the paper where the figure first appeared.	**Sections:** Paper progresses in a logical manner; providing background on the field, identifying the research question addressed by the paper, explaining the methods used to answer it, discussing the overall merit of the work in demonstrating their claims, and proposing steps for future work.

Figures: Paper contains at least one figure, with a caption. All figures are large enough to see.	**Paragraphs:** All information in each paragraph is clear, coherent, and related.
No plagiarism: Text is written in student's own words, including figure captions. Excessive similarity to another student's paper will be considered academic dishonesty, excessive similarity to published papers or online sources will be considered failure to summarize in original words and results in a revision.	**Transitions:** Each paragraph has a clear and coherent topic sentence that ties together the section in which it resides (i.e. each topic sentence transitions logically from the prior paragraph).
Summary: Does not include quotations, whether long blocks of text or multiple short phrases. These are not plagiarism, but they do not fulfil the requirements of the assignment.	**Technical Writing:** Student avoids overly wordy, confusing or "flowery" text. Sentences are straightforward; no run-ons.
Research Problem: Provides a clear statement of the scientific or technological research question the work addresses.	**Figures:** All figure(s) are referenced in the text to support a claim.
Background: Briefly describes the state of the field before the main paper to provide context for the current work.	**Terminology:** No errors in chemical biology terminology.
Methods: Briefly and clearly describes the experimental or theoretical methods used.	**Definitions:** Technical terms are defined, experiments not discussed in class are explained.
Discussion: Clearly explains at least one major experiment, simulation, or theoretical result from the paper. Explains the logic step by step and describes each result. Depending on the paper, more than one may be necessary to explain the take-home message.	**Methods:** Justifies the choice of which experiment(s) or simulation(s) are included.
Controls: Correctly identifies quality control metrics from the result described above. Not every paper has positive and negative controls, but all should have some type of quality control.	**Figures:** All figure(s) present in the paper are appropriate to illustrate important aspects of the main paper or background information.
Conclusions / Future Work: Student provides a reasonable next step for this line of research.	**Clarity:** Writing is clear and makes sense, without missing words, switches in tense, or other problems impacting understanding.
Grammar: The writing is grammatically correct such that it does not distract from the ideas presented (fewer than 2 unclear sentences).	**References:** 3-5 appropriate references and 0 inappropriate references are used.
High pass	All core and total ≥ 21
Low pass	All core and total ≥ 17
Needs Revision	Not all core met and /or total < 17

Table 2: Overall Grade Determination Matrix

Course Components	**Criteria Required to be Met to Earn Letter Grade**		
	A	B	C
Discussion Section Worksheet	9/10 complete	8/10 complete	7/10 complete

Problem Sets	4/5 high pass 1/5 low pass	3/5 high pass 2/5 low pass	1/5 high pass 4/5 low pass
Quizzes	3/4 high pass 1/4 low pass	2/4 high pass 2/4 low pass	1/4 high pass 3/4 low pass
Nucleic Acid Research Paper	high pass	1 high pass 1 low pass	low pass
Protein Research Paper	high pass		low pass

Table 3: Changes to Validated Survey Questions

Original Question[65]	Adjusted Question	Assessment Factor[65]
Q1 How confident are you that you could critique an experiment described in a **biology textbook** (i.e., list the strengths and weaknesses)?	Q1 How confident are you that you could critique an experiment described in a ***journal article*** (i.e., list the strengths and weaknesses)?	Methods of chemical biology
Q2 How confident are you that you will be successful in this **biology** course?	Q2 How confident are you that you will be successful in this ***chemical biology*** course?	Generalization to other chemical biology / science courses and analyzing data
Q3 How confident are you that you will be successful in **another biology** course?	Q3 How confident are you that you will be successful in ***a molecular biology*** course?	Generalization to other chemical biology / science courses and analyzing data
Q4 How confident are you that you will be successful in an **ecology** course?	Q4 How confident are you that you will be successful in an ***analytical chemistry*** course?	Generalization to other chemical biology / science courses and analyzing data
Q7 How confident are you that you could explain something that you learned in this **biology** course to another person?	Q7 How confident are you that you could explain something that you learned in this ***chemical biology*** course to another person?	Generalization to other chemical biology / science courses and analyzing data
Q8 How confident are you that after reading an article about a **biology** experiment, you could write a summary of its main points?	Q8 How confident are you that after reading an article about a ***chemical biology*** experiment, you could write a summary of its main points?	Application of chemical biology concepts and skills
Q9 How confident are you that after reading an article about a **biology** experiment, you could explain its main ideas to another person?	Q9 How confident are you that after reading an article about a ***chemical biology*** experiment, you could explain its main ideas to another	Application of chemical biology concepts and skills

	person?	
Q10 How confident are you that after watching a television documentary dealing with some aspect of **biology**, you could write a summary of its main points?	Q10 How confident are you that after watching a television documentary dealing with some aspect of ***chemical biology***, you could write a summary of its main points?	Application of chemical biology concepts and skills
Q11 How confident are you that after watching a television documentary dealing with some aspect of **biology**, you could explain its main ideas to another person?	Q11 How confident are you that after watching a television documentary dealing with some aspect of ***chemical biology***, you could explain its main ideas to another person?	Application of chemical biology concepts and skills
Q12 How confident are you that after listening to a public lecture regarding some **biology** topic, you could write a summary of its main points?	Q12 How confident are you that after listening to a public lecture regarding some ***chemical biology*** topic, you could write a summary of its main points?	Application of chemical biology concepts and skills
Q13 How confident are you that after listening to a public lecture regarding some **biology** topic, you could explain its main ideas to another person?	Q13 How confident are you that after listening to a public lecture regarding some ***chemical biology*** topic, you could explain its main ideas to another person?	Application of chemical biology concepts and skills
N/A	***Q14 How confident are you that you could apply concepts learned in this chemical biology course to a research project?***	Application of concepts to research project (*unvalidated addition)

Table 4: Breakdown of Token Usage

Approved Token Use	Total Number Used	% of Total
Missed Discussion Section	13	3.5
Problem Set (Late Submission)	28	7.4
Problem Set (Revision)	22	5.8
Quiz (Revision)	103	27.3
Quiz (Full Redo)	21	5.6
Writing Assignment (Late Submissions)	28	7.4
Writing Assignment 1 – Nucleic Acid Research Paper (Revision)	96	25.5
Writing Assignment 2 – Protein Research Paper (Revision)	66	17.5
Total	377	

Table 5: Writing Assignment "Needs Revision" Criteria

Criteria (*=core)	WA1	WA1 Resub	WA1 Resub 2	WA2	WA2 Resub	Δinitial WA1 and initial WA2	Δfinal WA1 Resub and final WA2 Resub
Page Limit *	4	3	0	6	3	2	0
Citations (format and placement) *	**25**	2	0	**35**	3	10	1
Figures (1 w/ caption, legible) *	14	2	1	13	0	-1	-2
No plagiarism *	1	0	1	6	0	5	0
Summary *	4	2	0	4	0	0	-2
Research Problem *	6	0	0	3	0	-3	0
Background *	18	1	0	13	0	-5	-1
Methods (describe) *	19	0	0	10	0	-9	0
Discussion *	**68**	3	0	**49**	3	**-19**	0
Controls *	**75**	12	0	**51**	7	**-24**	-5
Conclusions *	37	1	0	37	0	0	-1
Grammar	6	0	0	5	0	-1	0
Spelling	5	1	1	5	0	0	-1
Sections	2	0	0	7	0	5	0
Paragraphs	10	0	0	2	0	-8	0
Transitions	8	0	0	7	1	-1	1
Technical Writing	**32**	2	0	8	1	**-24**	-1
Figures (referenced in text)	4	2	0	7	1	3	-1
Terminology	16	1	1	18	4	2	3
Definitions	**29**	1	1	23	0	-6	-1
Methods (justify use)	9	1	0	6	1	-3	0
Figures (appropriate)	3	4	0	5	1	2	-3
Clarity	**37**	2	0	**36**	5	-1	3
References (3-5 appropriate)	8	2	0	6	0	-2	-2

Table 6: Student Grades on Writing Assignments

Assignment	Needs Revision	Low Pass	High Pass
Nucleic Acid Research Paper Initial	87	0	10
Nucleic Acid Research Paper Final	0	2	95
Protein Research Paper Initial	65	0	31
Protein Research Paper Final	0	7	90

Table 7: Comparison of Week 1 and Week 8 Responses to Survey Items

Survey Factor	Pre-Course Mean (sd)	Post-Course Mean (sd)	p-value
Methods of chemical biology	2.78 (0.80)	3.35 (0.89)	<0.001
Generalization to other chemical biology / science courses and analyzing data	3.26 (0.88)	3.51 (0.67)	0.004
Application of chemical biology concepts and skills	3.32 (0.87)	3.68 (0.76)	<0.001
Application of concepts to research project (**unvalidated addition*)	3.21 (1.04)	3.51 (1.00)	0.012