Specifications Grading in an Upper-Level BME Elective Course

Dr. Brian P. Helmke, University of Virginia

Brian Helmke is currently Associate Professor of Biomedical Engineering at the University of Virginia. He received the B.S.E. in bioengineering from the University of Pennsylvania, the B.S.Econ. from The Wharton School of the University of Pennsylvania, and the Ph.D. in bioengineering from the University of California, San Diego. Brian's research interests include cardiovascular physiology, cellular mechanobiology, and nanotechnology-based biomaterials. He is also interested in technology-enhanced teaching and in experiential learning for undergraduates in science and engineering.

Specifications Grading in an Upper-Level BME Elective Course

Recent trends in BME education emphasize aspects of the engineering profession such as design process, entrepreneurial mindset, and active problem-solving. However, the grading strategy in most traditional BME courses revolves around assigning points to student work based on apparent quality or degree of completion. Awarding “partial credit” is time-consuming and often is not closely mapped to learning objectives [1]. As a result, students often focus on how many points they earned relative to their perceived level of effort rather than on how closely they met the learning objectives of the activity or assessment.

In a “specifications grading” system [2], students earn credit for completing activities (or bundles of activities) by meeting clearly defined specifications shared at the time of assigning the activities. If the work does not meet the specifications, then credit is not earned. This system has several advantages. Specifications are closely mapped to the learning objectives for the activities and the course, making it easier to document and to reflect on learning. Students focus their effort on meeting specifications much as they would in the professional field when addressing client needs or competing for a project bid. Specifications can include aspects of the problem-solving or design process, which is often difficult to assess in a traditional points-based formula or rubric. Finally, the grading system is transparent, since students know up front what work must be completed to receive credit. Students can choose what grade level they wish to achieve, providing them with control over their learning and potentially increasing motivation [3].

A clearly defined set of specifications supports learning course content in depth, especially if the course design includes a variety of activities to increase student engagement [4]. However, it is not clear how students might perceive the learning environment with a specifications grading system. Resistance and frustration may arise from an active learning–based course design [5], especially for junior- and senior-level undergraduates who have become more established in their college learning habits. The emotional response associated with reduced learner satisfaction may inhibit the student’s ability to self-regulate effort towards learning and may therefore limit the gains achieved by implementing a goal-based grading system. Low-stakes formative assessments may be included to improve motivation, encourage risk-taking, and reduce anxiety [6].

This study aimed to test the hypothesis that a specifications grading system would increase student motivation and satisfaction by offering choice in learning activities while simultaneously increasing learning.

Research Methods

Course Design

The hypothesis was tested in BME 4641/ECE 4641 Bioelectricity, an elective course at the University of Virginia populated primarily by 3^{rd}- and 4^{th}-year undergraduate biomedical engineering and electrical engineering students. All course offerings included in the study were taught by the same instructor. The course content was the same in all course offerings and consisted of four units: (1) electrical properties of cell membranes, (2) the Hodgkin-Huxley model of action potential propagation, (3) synaptic transmission, and (4) measurement and analysis of bioelectric signals. The courses met twice each week for 75 min.

In a typical learning cycle surrounding a class session, students were assigned a textbook reading to complete before class. The class session consisted of alternating periods (averaging 10-15 min) of interactive lecture and student work time. Interactive lectures clarified and reinforced foundational knowledge and its application from the reading assignments, and students were expected to respond to questions about the reading, to fill in connections to previous class material, and/or to volunteer examples from their own experiences (in other classes, internships, research projects, etc.). Student work time enabled students to work on assignments. Since these work times were usually not long enough to allow for assignment completion, students were encouraged (but not required) to use this time to set up a solution framework. After class, students could complete the assignments they began in class and submit them electronically before the beginning of the next class. The beginning of the next class began with discussion of solutions from these assignments.

Five types of assignments were available to students: concept questions, practice problems, homework problems, unit tests, and an advanced project. Concept questions were designed to help students learn to connect detailed course content with their outside experiences, other courses in the curriculum, and their own career goals. Some of these questions asked students to reflect on and self-assess their own learning processes. Practice problems were similar to homework and test problems. The advanced project was a group project that involved visiting a lab to acquire EEG data during an "oddball" experiment and performing data reduction and analysis to identify key features of the EEG signal.

Points Grading System

As a control group, two sections of the course (fall 2016, 21 students; fall 2017, 30 students) were graded using a traditional points-based grading system. In order to give students choices of assignments to complete, two portfolios of assessments were created for each unit. A "traditional" portfolio consisted of homework problems (50% of the grade) and a unit test (50%). A "blended" portfolio was comprised of the same homework problems (35% of the grade) and unit test (35%), as well as concept questions (15%) and practice problems (15%). Concept questions and practice problems were low-stakes, "lightly graded" (for completion only) formative assessments. In the blended portfolio, the weight of the summative assessments (homework problems and unit tests) was adjusted to accommodate the low-stakes formative assessments. For the first unit of the course, students were required to complete the blended portfolio to expose them to the active learning–based style. This experience allowed them to make an informed choice of their preferred portfolio for Units 2, 3, and 4. A student choosing the traditional portfolio was instructed to complete only the homework and test for the unit. Students who were undecided were allowed to complete the low-stakes activities and choose their preferred portfolio based on the higher grade of the two. Since previous analysis demonstrated that summative assessment results and course grades were not different among students in the two portfolio groups [7], the results were combined for this study and termed, "Points Grading System."

Specifications Grading System

In the intervention group (fall 2018, 17 students), a specifications grading system was implemented (Appendix A). The work in the course was organized into "bundles" that reflected different levels of complexity when interacting with course content. Each bundle included concept questions, practice problems, homework problems, and unit tests. Completing an advanced project was required to complete the 'A' bundle. An activity was "completed" when the submitted work met all specifications for the activity, and specifications were designed to elicit a quality of work roughly equivalent to a 'B' score in a traditional grading system. Students earned a course grade by choosing to complete the minimum number of each type of activity in a bundle.

Since concept questions and practice problems were low-stakes activities designed to help students learn new concepts and connect them to their professional goals outside the class, specifications emphasized problem-solving process, risk taking, and identifying roadblocks to learning. A completely "correct" answer was not required. Homework and tests were high-stakes assessments in which students had the opportunity to demonstrate mastery of content. Specifications assessed problem-solving process, providing assumptions and evidence to support answers, and arriving at a substantially correct answer. The advanced project involved advanced lab and data analysis activities in an open-ended problem.

Since learning and mastering content is an iterative process, a currency of "tokens" was created to support students in their effort (Appendix B). Tokens were earned for completing reflection activities about individual learning and effort. Tokens could be redeemed for flexibility on assignment deadlines or for opportunities to revise and update submitted work.

Comparing Learning Outcomes

Course grades for the points-graded and the specifications-graded sections were compared directly as grade histograms. Since the minimum specifications for meeting learning objectives were designed to correspond roughly to a 'B', a "successful" score in the points-graded section was defined to be >85%. To compare scores on high-stakes activities (homework problems, unit tests, and advanced project), the proportion of students with successful scores (>85%) in the points-graded sections was compared to the proportion of students scored as "meets specifications". Since students in the specifications-graded section had the opportunity to revise and resubmit work that did not meet specifications on the first try, the proportion of students scored as "meets specifications" after a revision and reflection statement was also tabulated.

Student Satisfaction Survey

Both the points-graded and specifications-graded groups completed an end-of-course survey that contained Likert-type questions about the learning environment and contributions of class activities to learning, learner satisfaction, and the quality of faculty-student interactions. For Likert-type questions, student answers were encoded on a five-point scale as "strongly disagree" (1 point), "disagree" (2 points), "undecided" (3 points), "agree" (4 points), and "strongly agree" (5 points). In general, a mean score of at least 4.0 for a question was interpreted to mean that students viewed the topic of the question positively. Mean (± standard deviation) scores for the points-graded and specifications-graded sections were compared using an unpaired t-test, and scores were considered to be significantly different at a type I error rate of 0.05. Effect size of differences was estimated using Hedges' *g* to account for different sample sizes [8].

At the end of each survey section, students were asked an open-ended question to provide additional comments to evaluate the perceived effect of the grading system on their motivation and learning. Example answers representing the most frequent responses are reported.

Motivated Strategies for Learning Questionnaire (MSLQ)

At the end of the course, the specifications-graded group completed the Motivated Strategies for Learning Questionnaire (MSLQ) [9], which assesses college students' motivational orientations and their use of different learning strategies for a college course. The MSLQ is a validated tool based on a general cognitive view of motivation [10]. The tool measures three areas of motivation: value (intrinsic and extrinsic goal orientation, task value), expectancy (control beliefs about learning, self-efficacy), and affect (test anxiety). The learning strategies section measures cognitive (rehearsal, elaboration, organization, critical thinking), metacognitive (planning, monitoring, regulating), and resource management (time and studying environment, effort management, peer learning, help-seeking) strategies. Students responded to individual items on a 7-point Likert scale ranging from "not at all true of me" (1 point) to "very true of me" (7 points). The scale score was computed by computing the mean of the items making up the scale. Some items were negatively worded, so scores were reversed before computing the mean score for the scale.

General Self-Efficacy (GSE) Scale

In order to assess a general sense of perceived self-efficacy with respect to everyday life, the specifications-graded group completed the GSE scale [11] at the end of the course. The GSE consists of ten items with responses on a 4-point scale. A total score was computed as the mean of the scores of the ten individual items.

Results

Learning Outcomes

Course grade histograms were tabulated to compare the overall grade distribution for the two grading systems (Figure 1). The overall distributions were similar. Grades of 'B' or better were earned by 91% of students in points-graded sections and 94% of students in the specifications-graded section. The distribution of grades appeared more continuous for the points-graded sections because the underlying scores were continuous. By defining grades using counts of items meeting specificaions, the specifications-based grades fell into more discrete categories, i.e., most grades were whole letter grades rather than +/- grades.

Since high-stakes assessments (homework problems and unit tests) were scored differently between the two grading strategies, a method for comparison was defined using points-based scores >85% and scores of "meets specifications". In the points-graded sections, the proportion of students scoring >85% on homework varied slightly among individual assignments (Figure 2). Overall, 83% of students scored above this benchmark in total across all homework assignments. In the specifications-graded section, the proportion of students scored as "meets specifications" on their first submission was more variable among assignments. A trend emerged of increased success on first submission for later assignments. Overall, 60% of the initial submissions totaled across all assignments met the benchmark. Students in the specifications-graded section were able to revise their submissions based on feedback and reflect on their learning during the revisions. After allowing for revisions, 99% of submissions were scored as "meets specifications".

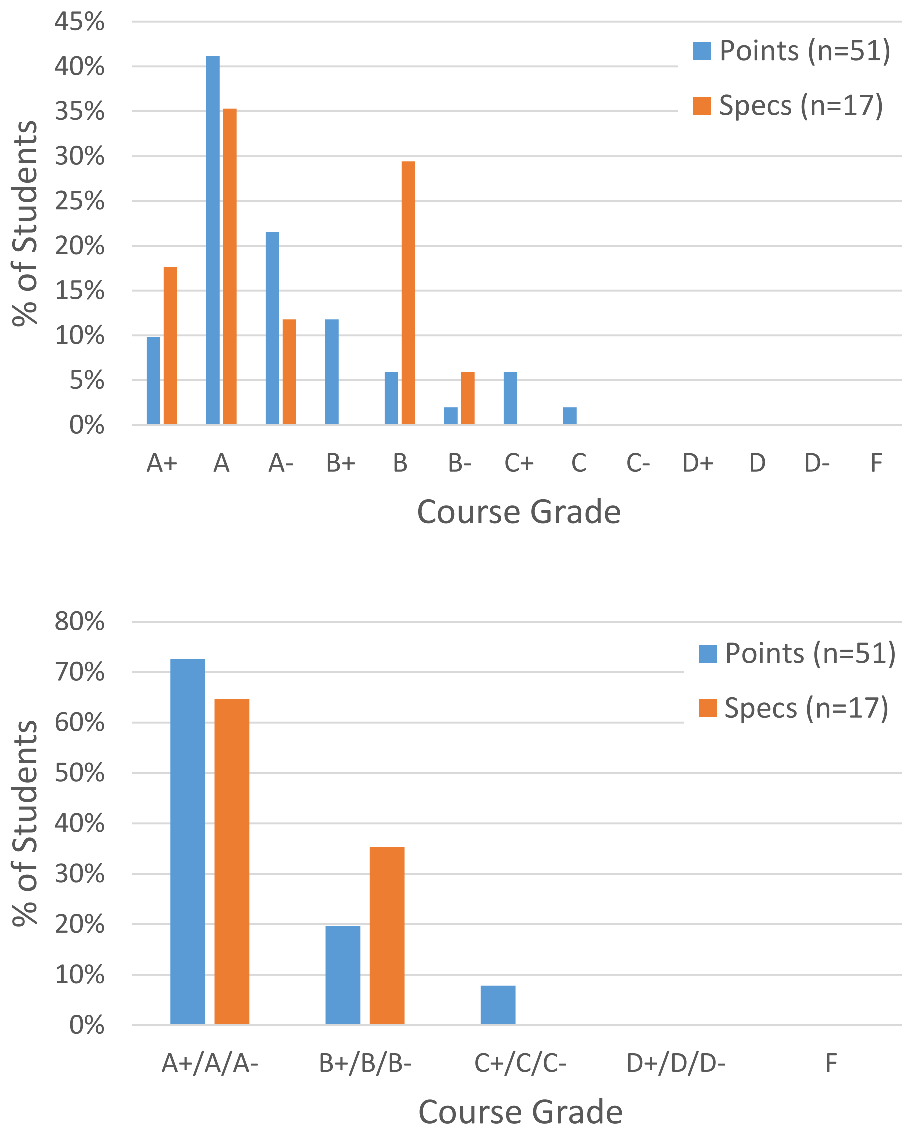

Figure 1. Final course grade histograms for point-graded (blue) and specifications-graded (orange) sections. Top figure shows grades binned with individual +/- grade categories, and bottom figure shows +/- grades binned together by letter category.

A similar comparison was performed for unit tests (Figure 3). In the points-graded sections, 65% of scores totaled over all tests were above 85%, although scores varied among the individual tests. In the specifications-graded section, fewer students were scored as "meets specifications" on their initial attempt for each test, and this trend was especially pronounced on the last test. Totaled across all tests during the semester, 25% were successful on first attempt. However, when students in the specifications-graded section were allowed second attempts at tests

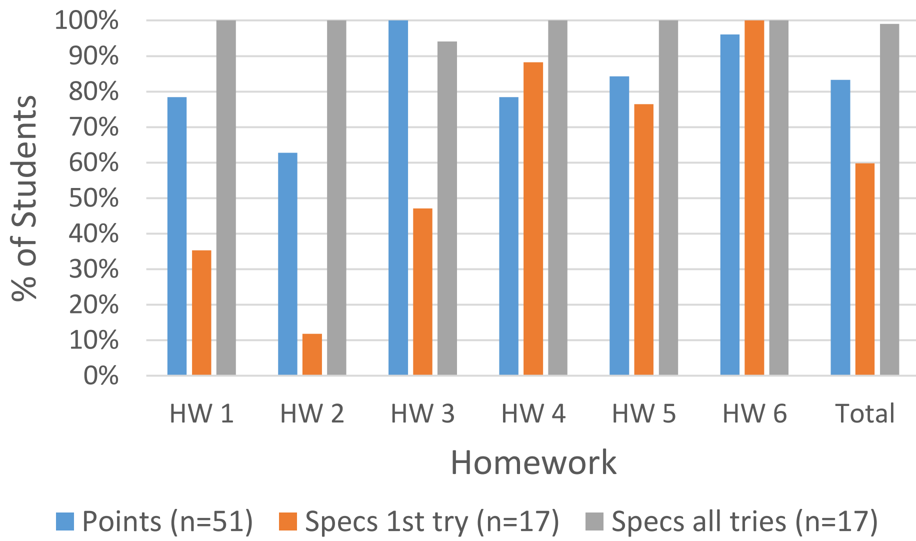

Figure 2. Proportion of students achieving homework scores >85% in the points-graded sections (blue) or scored as "meets specifications" on the first attempt (orange) or after one or two attempts (gray) in the specifications-graded section.

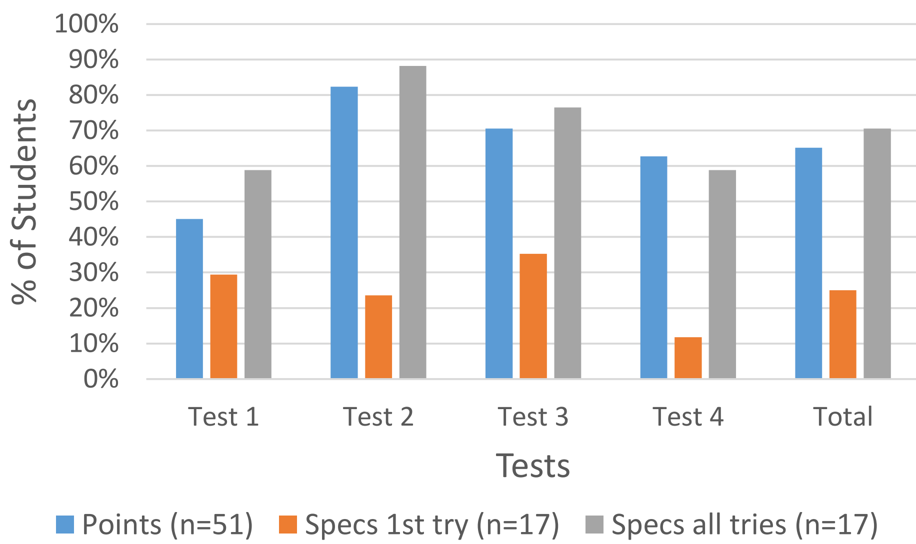

Figure 3. Proportion of students achieving unit test scores >85% in the points-graded sections (blue) or scored as "meets specifications" on the first attempt (orange) or after one or two attempts (gray) in the specifications-graded section.

Table 1. Learning Environment. Likert-type scores were encoded on a scale from 1 (strongly disagree) to 5 (strongly agree). Mean ± SD, p-value for unpaired t-test, Hedges' g value for effect size.

	Points (n = 51)	Specs (n = 13)	
Overall, the learning environment in this course was supportive and helped me learn.	4.35 ± 0.76	3.77 ± 0.80	$p = 0.02$ $g = 0.76$
The resources posted on the course website before class discussion helped me learn.	4.06 ± 0.78	4.00 ± 0.56	$p = 0.80$ $g = 0.08$
The class discussions helped me explore the class content.	4.29 ± 0.64	3.92 ± 83	$p = 0.09$ $g = 0.55$
The Concept Questions and Practice Problems helped me learn.	4.49 ± 0.64	4.23 ± 0.70	$p = 0.21$ $g = 0.40$
Homework problems and test questions helped me assess my progress learning the course content.	4.12 ± 0.62	3.69 ± 1.20	$p = 0.09$ $g = 0.56$
The structure of this course encouraged me to explore outside resources to help me learn.	3.94 ± 1.07	3.85 ± 0.77	$p = 0.77$ $g = 0.09$
I can relate what I learned in this course to other courses, my Capstone/Thesis project, and topics in the fields of biomedical engineering and medicine.	4.12 ± 0.88	4.31 ± 0.61	$p = 0.47$ $g = 0.23$
Overall average	4.20 ± 0.49	3.97 ± 0.52	$p = 0.15$ $g = 0.46$

containing similar problems, the success rate was similar to that in the points-graded sections. Overall, 71% of students met specifications totaled over all test attempts during the semester.

Learning Environment

Student perceptions of the learning environment were assessed using a series of seven Likert-type questions encoded on a five-point scale (Table 1). Although the overall average of the scores was not different between the two grading systems, students in the specifications-graded section responded more negatively to the overall supportiveness and helpfulness of the learning environment in the course. In the points-graded sections, the highest scores were associated with helpfulness of the low-stakes concept questions and practice problems, in-class discussions, and

Table 2. Learner Satisfaction. Likert-type scores were encoded on a scale from 1 (strongly disagree) to 5 (strongly agree). Mean ± SD, p-value for unpaired t-test, Hedges' g value for effect size.

	Points (n = 51)	Specs (n = 13)	
I liked the teaching style and learning environment in this course.	4.26 ± 0.74	3.08 ± 1.21	$p < 0.01$ $g = 1.39$
I am satisfied with how well/how much I learned in this course relative to my level of effort towards learning.	4.20 ± 0.77	3.69 ± 1.20	$p = 0.07$ $g = 0.58$
I am satisfied with the number and quality of opportunities to assess my own understanding and learning that I received in this course.	4.29 ± 0.64	3.46 ± 0.93	$p < 0.01$ $g = 1.19$
Overall average	4.25 ± 0.64	3.41 ± 0.97	$p < 0.01$ $g = 1.18$

the overall supportiveness of the learning environment. In the specifications-graded section, the highest scores were associated with helpfulness of the low-stakes concept questions and practice problems and with ability to relate learning in this course to other courses, Capstone/Thesis projects, and topics in the fields of biomedical engineering and medicine. The lowest score in the specifications-graded section was the helpfulness of homework problems and test questions in assessing progress learning the course content.

Learner Satisfaction

Students responded to three Likert-type questions about their satisfaction with their own learning in the course (Table 2). The overall score for this construct was significantly lower (with very large effect size) in the specifications-graded section than in the points-graded sections. The lowest score was associated with liking the teaching style and learning environment in the course, and none of the individual item scores were >4.0, which is often associated with a "positive" response in a 5-point Likert scale.

Student-Faculty Interactions

Students responded to five Likert-type questions encoded on a five-point scale about the quality of student-faculty interactions in the course (Table 3). The overall average score for this construct was significantly lower (with medium-large effect size) in the specifications-graded section than in the points-graded sections. The biggest difference in scores was for the item associated with level and quality of student-instructor interactions in the course ($p < 0.01$, very large effect size). The lowest score in the specifications-graded section was associated with the

Table 3. Student-Faculty Interactions. Likert-type scores were encoded on a scale from 1 (strongly disagree) to 5 (strongly agree). Mean ± SD, p-value for unpaired t-test, Hedges' g value for effect size.

	Points (n = 51)	Specs (n = 13)	
The structure of this course encouraged me to interact with my instructor and teaching assistant.	4.10 ± 0.87	3.77 ± 0.58	$p = 0.21$ $g = 0.40$
I am satisfied with the level and quality of student-instructor interactions in this course.	4.49 ± 0.57	3.85 ± 0.66	$p < 0.01$ $g = 1.09$
Student-instructor interactions in this course were supportive and helped me learn.	4.39 ± 0.72	4.15 ± 0.53	$p = 0.27$ $g = 0.35$
The instructor was readily accessible when I needed help with my learning.	4.51 ± 0.67	4.23 ± 0.58	$p = 0.18$ $g = 0.43$
I am satisfied with the amount and quality of feedback about my progress toward course objectives that I received in this course.	3.94 ± 0.98	3.30 ± 1.26	$p = 0.06$ $g = 0.61$
Overall average	4.29 ± 0.61	3.86 ± 0.47	$p = 0.02$ $g = 0.73$

amount a quality of feedback about progress toward course objectives. More positive scores >4.0 were recorded for items associated with supportiveness and helpfulness of student-instructor interactions and with accessibility of the instructor.

Open-Ended Responses: Grading Strategy and Student Motivation

In the points-graded sections, students were asked to respond to the following open-ended prompt: "How did having a choice of grading strategy affect your motivation and learning in this course?" Out of 51 responders, 15 (29%) reported that the choice of grading strategy increased their motivation and incentivized and/or improved their learning in the course. Twenty-two students (43%) reported that the grading strategy had no effect on their motivation, and 14 (27%) of these students indicated that they planned from the beginning of the semester to complete all assignments independently of the grading strategy. Two students (4%) reported that the grading strategy reduced their motivation. Both students gave reasons associated with a desire to complete only the minimum possible work to achieve their desired grade in the course. Students who experienced increased or no effect on motivation often commented that the grading strategy reduced their stress associated with grading (15 responses, 29%) so that they could focus on their learning and on gaining a conceptual understanding of course material (5 responses, 10%). Two students (4%) expressed appreciation that the instructor cared about them individually enough to accommodate their learning style.

In the specifications-graded section, students were asked to respond to the following prompt: “How did having choices in the grading strategy affect your motivation and learning in this course?” Twelve students responded. Two students felt that the grading strategy increased their motivation to learn because of increased schedule flexibility, ability to resubmit revised work, and clear expectations. Five students remarked that the grading strategy decreased their motivation and learning because of increased anxiety and stress. These students perceived the stakes were higher for each assignment or activity, believed the expectation was that all work had to be completed perfectly in order to earn credit, and felt uncertainty near the end of the semester waiting for scores and feedback on the last assignments to be returned.

Both the points-graded and specifications-graded sections were asked to respond to the open-ended prompt, “Comment on how student-instructor interactions affected your learning in this course.” In the points-graded sections (51 responses), students commented most frequently about the availability (15 mentions) and approachability (7 mentions) of the instructor and the supportiveness of office hours (8 mentions), in-class interactions (8 mentions), and electronic responses via email or Piazza (8 mentions). Four students commented that feedback on assignments was delayed. In the specifications-graded section (13 responses), students mentioned most frequently positive interactions with the instructor in office hours (4 mentions) and in class (4 mentions). Two students commented that feedback on assignments was delayed.

Motivated Strategies for Learning Questionnaire (MSLQ)

The MSLQ was administered only to the specifications-graded course, so comparisons to the points-graded courses were not available. Instead, a score significantly greater than 5.0 was considered a “positive” response, and data were compared using a one-sample t-test ($p < 0.05$). In the motivation section of the MSLQ (Table 4), none of the components were scored positively. The component with the highest score was the Control of Learning Beliefs scale, which refers to the belief that one’s efforts to learn will result in positive outcomes.

In the learning strategies section of the MSLQ, two components scored positively: Elaboration, and Effort Regulation. Elaboration strategies help one store information into long-term memory by creating connections among new items to be learned. Effort regulation is an aspect of self-regulation associated with an ability to control one’s effort when faced with distractions and uninteresting tasks. In this section, critical thinking, self-regulation, and peer learning components appeared to score low.

General Self-Efficacy (GSE) Scale

The GSE Scale (Table 5) was administered only to the specifications-graded course, so comparison to the points-graded courses was not available. The overall average score was 3.17 ± 0.36 (mean ± SD), not significantly greater than 3.0 (one-sample t-test, $p < 0.05$).

Table 4. Motivated Strategies for Learning Questionnaire (MSLQ), Motivation section. Likert-type scores were encoded on a scale from 1 (not at all true of me) to 7 (exactly true of me). For each construct, the overall average of individual item scores is reported as mean ± SD.

Value Component: Intrinsic Goal Orientation (n = 13)	4.96 ± 1.18
Value Component: Extrinsic Goal Orientation (n = 13)	4.67 ± 0.96
Value Component: Task Value (n = 11)	5.02 ± 1.09
Expectancy Component: Control of Learning Beliefs (n = 13)	5.40 ± 0.79
Expectancy Component: Self-Efficacy for Learning and Performance (n = 12)	4.70 ± 1.29
Affective Component: Test Anxiety (n = 13)	4.58 ± 1.67

*Table 5. Motivated Strategies for Learning Questionnaire (MSLQ), Learning Strategies section. Likert-type scores were encoded on a scale from 1 (not at all true of me) to 7 (exactly true of me). For each construct, the overall average of individual item scores is reported as mean ± SD. *p < 0.05, one-tailed t-test, score >5.0.*

Cognitive and Metacognitive Strategies: Rehearsal (n = 13)	4.85 ± 1.26
Cognitive and Metacognitive Strategies: Elaboration (n = 13)	5.46 ± 0.51*
Cognitive and Metacognitive Strategies: Organization (n = 13)	5.12 ± 0.84
Cognitive and Metacognitive Strategies: Critical Thinking (n = 13)	4.28 ± 1.37
Cognitive and Metacognitive Strategies: Metacognitive Self-Regulation (n = 13)	4.69 ± 0.41
Resource Management Strategies: Time and Study Environment (n = 13)	5.38 ± 1.16
Resource Management Strategies: Effort Regulation (n = 13)	5.50 ± 0.95*
Resource Management Strategies: Peer Learning (n = 12)	4.00 ± 1.84
Resource Management Strategies: Help Seeking (n = 13)	4.63 ± 1.25

Discussion

In this project, the goal was to redesign Bioelectricity around a “specifications grading” paradigm [1, 2]. This pedagogical approach involves creating activities that are evaluated based on a rubric for successful completion of the activity, similar to a “pass/fail” grade. The specs are derived directly from the course learning objectives; they are designed to help students achieve competency in these objectives. Specs are designed for each activity and assessment, whether formative or summative. The advantage of this approach is that expectations for learning are clearly communicated to students at the beginning of the activities. Students are granted agency to choose which activities they believe will benefit them according to their priorities for learning, but they must choose to complete activities to fulfill all of the course objectives in order to pass the course. They receive credit for an activity only if they meet all of the specs defined for that activity. It is important to note that this type of course design can improve accommodation of students with differing backgrounds and abilities by outlining specs in an inclusive manner and pairing with inclusive modes of content delivery.

The proportion of students earning a course grade of ‘B’ or better was similar in the specifications-graded section and in the points-graded sections. In the specifications-graded section, all students met specifications on essentially all homework assignments. Since the homework assignments spanned all of the course learning objectives, this result implies that all students met the primary learning objectives for the course. Differentiating student performance based on level of depth or expertise in some content areas was achieved using unit tests. The course grade was linked to the number of content areas (unit tests) in which the students met specifications. In addition, a student was required to meet specifications for an advanced project to earn an ‘A’ in the course. Thus, students who earned an ‘A’ demonstrated mastery in at least three of four major content areas. The specifications grading system also provided flexibility because students were not required to meet specifications on all unit tests. Students could choose to focus their efforts in content areas of most interest to them.

In contrast, under the points-graded system, it is difficult to say exactly how many course objectives were fulfilled by each student, since the grade is determined by earning “enough” points across a number of assignments. A student earning a ‘B’ may have fulfilled parts of several learning objectives and fulfilled none of the objectives completely. Alternatively, that student may have earned more points for in-depth understanding of some learning objectives while earning few points and not meeting other learning objectives. In both examples, the grade outcome is the same. Consequently, it is more difficult to assess student learning using grades, which may have an impact on the ability to map student learning to objectives for standardization such as in accreditation criteria.

An important aspect of the specifications grading system is the ability of students to revise and resubmit their work. In this course, a token currency was created to provide second chances and to motivate reflection on learning. Students earned tokens for completing test wrappers and answering reflection questions about the learning strategies in the course. These activities were designed to stimulate self-regulated learning, which is the ability to plan, monitor, and evaluate one’s learning [12]. After earning tokens, students could spend them to extend deadlines, to meet specifications on low-stakes assignments, to revise and resubmit a homework, or to retry a test.

The goal was to reduce anxiety and increase agency by providing flexibility in meeting the course learning objectives.

Low-stakes activities were designed to introduce students to new topics, to promote appreciation of both historical and contemporary experimental and modeling approaches, and to connect course content to students' own experiences and career plans. These assignments encouraged students to take risks by outlining problem solutions even if they were unsure of the validity of some of their answers. This approach also helped students identify, evaluate, and defend assumptions in their problem-solving approaches. In the points-graded sections, students earned full points for a good faith effort, and their appreciation for this approach was demonstrated in their responses to the open-ended survey question. Several students commented about the satisfaction of feeling comfortable enough to try out a problem solution without the associated stress or worry of getting the answer perfectly correct. Interestingly, the students' feedback in the specifications-graded section was mixed. For example, one student perceived that a perfectly correct answer was required in order to meet specifications, even though the specifications were defined based on good faith effort and a learning reflection statement. Other students appreciated the clarity of expectations outlined with each assignment and felt that the ability to do revisions and resubmissions reduced anxiety. It is not clear why students in the points-graded course felt that these activities were helpful while students in the specifications-graded course had anxiety about not earning credit. One possibility is that increased anxiety associated with high-stakes activities in the specifications-graded course was transferred to these activities. In this case, the increased anxiety would inhibit students from taking risks and exploring the course content.

Even though the grade distribution in the specifications-graded section was similar to that in the points-graded sections, student satisfaction with the learning environments differed. Choices of low-stakes learning activities in the points-graded sections resulted in a positive student perception of the overall learning environment, satisfaction with learning relative to effort, and appreciation for the ability to skip some formative activities if students were not as interested in the topic or were stressed about other time commitments. This level of satisfaction may contribute to reduced student resistance and frustration in active-learning based courses [5].

In the specifications-graded section, the average score for the overall learning environment was significantly lower than for the points-graded section. The biggest differences were associated with class discussions, homework problems, and test questions (Table 1). The large standard deviations and student comments about the grading system conveyed variability in student perceptions of the learning environment. Some students felt increased anxiety because they believed that perfection was required on every assignment to earn credit. This self-induced pressure on every assignment was increased near the end of the semester. One student believed that completing all activities was required to earn an 'A', which would mean students didn't actually have choices. Despite these concerns, one student admitted that maybe they just weren't used to the system yet, as it seemed to be getting easier later in the semester. Although other courses at the institution use a specifications grading system (including in computer science), students in this course had not experienced those courses. Since this was students' first experience with this type of grading system, they were unsure how to strategize their efforts towards grading. Newness of the grading system would disrupt student habits, increase resistance, and decrease comfort level. These responses may partly explain increased anxiety and reduced satisfaction with the learning environment.

In contrast, other students shared comments reflecting they were more satisfied with the specifications grading system. One student commented that specifications grading helped to focus on learning content areas rather than grading outcomes because full credit was earned by meeting the specifications. Another student mentioned that flexible scheduling with submissions and the ability to follow up with revisions increased satisfaction because they were balancing time commitments required for job interviews. These positive comments indicate a bimodal response from groups of students that is not reflected in the lower mean score associated with the learning environment. Other studies have reflected a similar phenomenon when students separate into groups that are either supportive of or resistant to an active learning course design [13].

Although the course style influences the quality of student-faculty interactions [14], the choice of grading system was not expected to have an impact. The course design in both cases was based on a mix of mini-lectures and active learning activities, which should have enabled frequent discussions about course content as well as advising or mentoring. Increased frequency and quality of student-faculty interactions support students' desire to connect their learning to their career development [15]. In fact, student-faculty interactions were viewed positively by students in the points-graded sections. Students especially appreciated the availability and approachability of the instructor in class, in office hours, and electronically. One example appears in the following comment:

> "[The instructor] was really supportive throughout this entire course. I really appreciated the fact that [the instructor] took the time at the very beginning of the course to ensure that I understood the basic concepts of membrane potentials even when that required to go step-by-step through the process. Personally, this set the tone for the rest of the class because I was able to go into this material knowing that there was a professor to support me through it. This really helped my attitude when approaching the material which helped me to learn more efficiently."

Surprisingly, the level and quality of student-instructor interactions were scored significantly lower in the specifications-graded section than in the points-graded sections. Interestingly, the responses to open-ended questions appear to disagree somewhat with the mean Likert score in the specifications-graded section. A number of students mentioned in response to the open-ended question they had positive interactions in class, in office hours, and electronically. One student commented, "[The instructor] was very interactive during lecture, it was very helpful to have [them] take our feedback and readjust the class structure, ask for our opinions about deadlines. It made things alot [sic] less stressful…." The decreased Likert score may have reflected negative perceptions associated with the grading system projected onto the instructor. For example, students expressing discomfort or dissatisfaction with the system harbored frustration and resisted interacting with the instructor in a positive manner. One such example of mixed signals appeared in the following comment:

> "I only attended office hours a few times, as majority of them were during other class periods. Most of them were helpful, especially when I was asking a question specific to an assignment or exam. When I expressed stress about this class and fear of doing poorly based on the grading structure, I felt my opinions were heard but opposed and not much changed. While I understand grading structures cannot

> be changed mid-way through the course, I constantly felt frustrated by this class and the way my grade reflected what I was actually learned."

The frustration expressed by this student probably impacted their motivation to learn in the course. The MSLQ is a validated instrument that measures value, expectancy, and test anxiety as components of motivation [9]. Test anxiety was not in the upper range of the Likert scale (>5.0), as might have been expected due to the opportunities to retake tests that did not meet specifications yet. However, value and expectancy were also not in the upper range. This result was surprising, since increased transparency and autonomy along with opportunities to connect to out-of-course activities were expected to enhance these aspects of motivation [3]. Interestingly, elaboration as a learning strategy and effort regulation as a resource management strategy scored highly (Table 5). Elaboration involves the individual activity of making meaning of new content to help commit it to long-term memory, and effort regulation is reflected by an ability to resist distractions and focus on learning tasks. One might have expected that these strategies would have been associated with high scores for the task value component of motivation. It is likely that these scores reflect assignments that were authentic to students' career interests, making it easier to draw connections to research projects such as Capstone or independent student projects.

Interestingly, critical thinking and peer learning strategies were attributes with the lowest scores. The critical thinking score was surprising because many assignments were designed to support learning problem-solving process, for example, applied to membrane electrophysiology or electrode design. Moreover, the active learning approaches used in class and in the advanced project were designed in part as peer teaching and peer learning approaches. The connection between activity design and these learning strategies will be explored in a future study.

Giving students the agency to decide which grade bundle to pursue was designed to support self-determination and self-efficacy [16]. In the points-graded sections, one students mentioned discovering that they enjoyed the increased autonomy in choosing activities to complete. In the specifications-graded section, two students commented that the ability to do revisions using the token system reduced anxiety. In both cases, the students were reporting reduced stress associated with grading.

Overall, this study described applying a specifications grading system to an upper level elective BME course. A number of benefits of the grading system were hypothesized. Course grades and summative assessment outcomes were similar when compared to the same course taught previously using a points-based grading scheme. However, student satisfaction with the learning environment was decreased in the specifications-graded course. Some aspects of student-faculty interactions associated with the grading scheme were less satisfactory, even though personal interactions supported learning equally well.

References

1. L. B. Nilson, "Yes, Virginia, there's a better way to grade," *Inside Higher Ed*, Jan. 19, 2016, https://www.insidehighered.com/views/2016/01/19/new-ways-grade-more-effectively-essay, retrieved Feb. 4, 2019.

2. L. B. Nilson, *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Sterling, VA: Stylus, 2015.

3. T. Garcia and P. R. Pintrich, "The effects of autonomy on motivation and performance in the college classroom," *Contemp. Educ. Psychol.*, vol. 21, pp. 477-486, 1996.

4. S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, "Active learning increases student performance in science, engineering, and mathematics," *Proc. Natl Acad. Sci. USA*, vol. 111, pp. 8410-8415, 2014.

5. D. U. Silverthorn, "Teaching and learning in the interactive classroom," *Adv. Physiol. Educ.*, vol. 30, pp. 135-140, 2006.

6. K. M. Cauley and J. H. McMillan, "Formative assessment techniques to support student motivation and achievement," *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, vol. 83, pp. 1-6, 2010.

7. B. P. Helmke, "Student choice of traditional or blended learning activities improves satisfaction and learning outcome," ASEE Annual Conference, June 24-27, 2018, Salt Lake City, UT. Paper ID #22763.

8. L. V. Hedges "Distribution theory for Glass' estimator of effect size and related estimators," *J. Educ. Stat.*, vol. 6, pp. 107–128, 1981.

9. P. R. Pintrich, D. A. Smith, T. Garcia, and W. J. Mckeachie, "Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ)," *Educ. Psychol. Meas.*, vol. 53, pp. 801-813, 1993.

10. W. J. McKeachie, P. R. Pintrich, Y. G. Lin, and D. Smith, D. *Teaching and learning in the college classroom: A review of the resecach literature*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, The University of Michigan, 1986.

11. R. Schwarzer and M. Jerusalem, M, "Generalized Self-Efficacy scale," in J. Weinman, S. Wright, and M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs.* Windsor, UK: NFER-NELSON, pp. 35-37, 1995.

12. B. J. Zimmerman, "Self-regulated learning and academic achievement: An overview," *Educational Psychologist*, vol. 25, pp. 3-17, 1990.

13. B. P. Helmke, "Barriers to learning in a large flipped biotransport course," ASEE Annual Conference, June 25-28, 2017, Columbus, OH. Paper ID #18299.

14. J. J. Endo and R. L. Harpel, “The effect of student-faculty interaction on students’ education outcomes,” *Res. Higher Educ.*, vol. 16, pp. 115-138, 1982

15. B. Christe, “The importance of faculty-student connections in STEM disciplines: A literature review,” J. STEM Educ. Innov. Res., vol. 14, pp. 22-26, 2013.

16. C. F. Brooks and S. L. Young, “Are choice-making opportunities needed in the classroom? Using self-determination theory to consider student motivation and learner empowerment,” *Intl J. Teaching Learning Higher Educ.*, vol. 23, vol. 48-59, 2011.

APPENDIX A. Sections of the course syllabus that describe the specifications grading system.

How will we evaluate your learning?

Grading creates stress. I wish it didn't exist. I would rather we focus on evaluating helping you achieve your learning objectives for this course. Your personal goal might be to achieve all of the objectives by the end of the semester, or it might be to gain familiarity with course topics without worrying about being perfect in everything. Either way is OK. My goal is to create a safe yet challenging learning environment that de-emphasizes grading and helps you reach your personal learning goal. We will do this using a system called "specifications grading". Under this system, you take ownership of your grade by deciding the amount and quality of work to do. Your grade is tied directly to expectations and learning. The work in the course is organized into bundles that reflect different levels of complexity, and you earn a grade by choosing which bundles to complete.

Activities within each bundle are closely mapped to learning objectives. You demonstrate your learning by completing an activity so that it "meets specifications". Specifications are the success criteria for each activity. Your work must satisfy all the criteria to earn credit for the activity. This system avoids the worry about how or why you earned partial credit or lost points. Instead, it mimics common practice in the professional world. For example, you either pass the Ph.D. Candidacy exam or you don't; you pass the P.E. Exam or you don't; you pass the Medical Board Exam or you don't; you meet expectations in your annual review with your boss or you don't. You get the idea.

The table below outlines how the assignment bundles are organized. You earn a grade by completing the minimum number of each assignment type in the bundle. I will share more details about the EEG Project during the first two weeks of class. For more details about the assignment types, click on them in the table. For examples of how grading works, click here.

Assignment	'A' bundle	'B' bundle	'C' bundle	'D' bundle
Concept Questions (12)	11	10	9	8
Practice Problem (12)	11	10	9	8
Homework (6)	6	6	5	4
Unit Tests (4)	3	2	1	0
EEG Project	X			

Notice that this system is not named "pass/fail grading". It doesn't help you learn if the activity ends when you "fail". It's OK if an aspect of your work does not meet specifications yet (just like in the professional world). Sometimes you have to revisit an idea multiple times or at a slower pace before you "get it". I want to encourage you to see learning as an iterative process. To support you in this effort, I have created a currency of "tokens". You may earn up to 5 tokens during the semester by completing certain activities that ask you to reflect on your learning. You may redeem tokens for assignment flexibility that will help you learn. For details about the tokens, click here.

Grading Examples

In this grading system, you choose the grading bundle that you will complete, so the course grade you earn is entirely up to you. For example, to earn an "A", you must meet specifications for 11 Concept Questions, 11 Practice Problems assignments, 6 Homeworks, 3 Tests, and the EEG Project. To earn a "B", you must meet specifications for 10 Concept Questions, 10 Practice Problems, assignments, 6 Homeworks, and 2 Tests; the EEG Project is not required for a "B".

Assignment	'A' bundle	'B' bundle	'C' bundle	'D' bundle
Concept Questions (12)	11	10	9	8
Practice Problem (12)	11	10	9	8
Homework (6)	6	6	5	4
Unit Tests (4)	3	2	1	
EEG Project	X			

You may earn a "Plus" grade by completing one additional Unit Test beyond the required number for your grade bundle. For example, you earn a "B+" if you complete 10 Concept Questions, 10 Practice Problems, 6 Homeworks, and 3 Tests (one more than required for a "B").

You may earn a "Minus" grade if you are only missing one Concept Question or one Practice Problem to complete a bundle. For example, You earn a "B-" if you complete 9 Concept Questions (one less than required for a "B"), 10 Practice Problems, 6 Homeworks, and 2 Tests.

Scenario 1. Jill met specifications for 11 Concept Questions, 10 Practice Problems, 6 Homeworks, and 2 Tests. She did not choose to complete the EEG Project.

Assignment	'A' bundle	'B' bundle	'C' bundle	'D' bundle
Concept Questions (12)	11	10	9	8
Practice Problem (12)	11	10	9	8
Homework (6)	6	6	5	4
Unit Tests (4)	3	2	1	
EEG Project	X			

Jill earned a "B" because she met specifications for all items in the "B" bundle.

Scenario 2. DeSean met specifications for 11 Concept Questions, all 12 Practice Problems, all 6 Homeworks, 4 Tests, and the EEG project.

Assignment	'A' bundle	'B' bundle	'C' bundle	'D' bundle
Concept Questions (12)	11	10	9	8
Practice Problem (12)	11	10	9	8
Homework (6)	6	6	5	4
Unit Tests (4)	3	2	1	
EEG Project	X			

DeSean earned an "A+" because they met specifications for all items in the "A" bundle and completed one extra Test.

Scenario 3. Manikandon met specifications for 12 Concept Questions, 10 Practice Problems, 6 Homeworks, 3 Tests, and the EEG Project.

Assignment	'A' bundle	'B' bundle	'C' bundle	'D' bundle
Concept Questions (12)	11	10	9	8
Practice Problem (12)	11	10	9	8
Homework (6)	6	6	5	4
Unit Tests (4)	3	2	1	
EEG Project	X			

Manikandon earned an "A-". He met specifications for the "A" bundle except for one Practice Problems assignment.

Concept Questions and Practice Problems

Concept Questions. Concept Questions ask for qualitative answers based on course readings and outside resources to help you describe biophysical mechanisms and to relate bioelectrical phenomena to your interests outside of this course. Questions may be completed in class or on your own, as specified by each assignment. I expect that you will spend <30 min on each Concepts Questions assignment.

Practice Problems. Practice problems are short quantitative problems designed to introduce tools you may use in more complex analyses to solve Homework and Test problems later. The emphasis is on practice, since the specifications for earning credit do not require a correct answer (see below). I expect that you will spend <60 min on each Practice Problems assignment.

Specifications. These assignments are graded either "Meets Specifications" (S) or "Incomplete" (I). You earn full credit when you meet all of the following specifications for all of the questions and problems in the assignment:

- Your logic and train of thought are organized and clearly presented in acceptable scientific communication style;
- You incorporate evidence such as class readings, outside resources, and prior knowledge to support your answers;
- An expert reader (usually Mike or me, but also your peers) can evaluate whether your answers and/or solution methods are correct; and
- You meet other specifications outlined in the individual assignment.

Notice that it's OK if you don't get a correct final answer! The specifications encourage you give your best effort and to make your learning process transparent so I can help you fill in the gaps.

Homework and Tests

Homework. By completing Homework problems, you meet the core learning objectives of the course. Problems are designed to demonstrate your quantitative reasoning skills and understanding of biophysical mechanisms.

Unit Tests. By completing Tests, you demonstrate mastery of the course material. The unit tests require you to apply and integrate course material to solve new problems. Tests will include both qualitative reasoning and quantitative problem solving.

Specifications. Homework and Tests are graded either "Meets Specifications" (S), "Progressing" (P), or "Incomplete" (I). You earn credit and a grade of "MS" by meeting <u>all</u> of the following specifications:

- Your logic and train of thought are organized and clearly presented;
- Your problem setup includes valid assumptions and shows an appropriate quantitative framework;
- Your problem solution methods are complete and substantially correct; and
- You meet other specifications outlined in the individual assignment.

In general, you will earn a grade of "S" even if there are a small number of trivial errors that do not cast doubt on your ability to meet the learning objective. A larger number of minor errors that can easily be corrected in a Homework revision or a Test retake will earn a grade of "P".

APPENDIX B

The Token Currency

To support you in your learning, I have created a currency of "tokens". You may earn up to 5 tokens during the semester by meeting specifications for the following activities that ask you to reflect on your learning:

- Self-evaluation of learning surveys (up to 2 tokens); and
- Test wrappers (up to 3 tokens), in which you self-evaluate your learning process.

To provide you with flexibility in your learning, you may redeem tokens for any of the following:

- Retry a Test that is "Incomplete" for a second chance to meet specifications. You may take the Test during any test period or office hours.
- Revise and resubmit a Homework that is "Incomplete", explaining your revisions and what you learned by revising your answers (due 48 hr after the original Homework is returned).
- Revise and resubmit a Concept Question or Practice Problem assignment that is "Incomplete", explaining your revisions and what you learned by revising (due 24 hr after the original assignment is returned).
- Extend a Homework deadline by 48 hr (must be requested at least 24 hr before the deadline).
- Automatically count one Concept Question or Practice Problem assignment as meets expectations (must be requested by midnight of the day the assignment is due).

You may only redeem one token per assignment; for example, you cannot redeem a token to extend a Homework deadline and also spend a token to revise and resubmit it. You do not need to redeem a token to revise a Homework or retry a Test that is marked "Progressing".