ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #37479

Turns Out Our Exams Were Pointless, So We Changed Our Assessment Strategy

Laura K Alford (Lecturer and Research Investigator)

I am a lecturer at the University of Michigan. I research ways to use data-informed analysis of students' performance and perceptions of classroom environment to support DEI-based curricula improvements.

Heather Rypkema

Harsh Manoj Jhaveri (Graduate Student Instructor, Master's Student)

Harsh Jhaveri is a master's student at the University of Michigan, pursuing a degree in Robotics. Harsh previously has completed dual bachelor's degrees in Aerospace Engineering and Computer Science Engineering, also from the University of Michigan. In addition to his pursuing his degree, Harsh is also a graduate student instructor (GSI) for Engineering 101, Introduction to Computers and Programming, a first-semester course mandatory for all engineering students. In addition to his teaching duties, Harsh has helped facilitate and develop course logistics, course development, and professional development for staff members through the Foundational Course Initiative at the University of Michigan. Outside of teaching, Harsh enjoys developing software for autonomous aircraft systems, cooking, and collecting Vinyl LPs.

Ryien Hosseini

Megan Beemer

Turns Out Our Exams Were Pointless, So We Changed Our Assessment Strategy

Abstract

This research paper describes our analysis of how student exam scores in a large introductory programming course evaluate student learning in the context of other assessment mechanisms. Data from Academic Years 2018-2019 and 2020-2021 were used to compare the pre-pandemic individual assessment scheme with the revised scheme implemented in response to the shift to remote instruction. Our analysis focused on two key questions: 1) How well are individual assessments enabling students to demonstrate their learning? and 2) how equitable are the assessments with respect to grade outcomes for students with historically marginalized identities? Specifically, our aim was to reduce the onerous workload of exam preparation for the instructional team, while assessing student knowledge in a pedagogically effective and equitable way.

Based on the results of this development and analysis, we implemented a four-assessment structure for Fall 2021. Preliminary analysis of these assessments indicate that the new assessments are more equitable and enable earlier identification of students who may be struggling with the course material. Additionally, the new assessment infrastructure requires significantly less instructor time to maintain and implement from term to term. In this paper, we will describe our motivation for this study, the analysis of past exam and course data, our thought process for the new assessments, and a preliminary analysis of the effectiveness and equity of the new assessments. We hope that others find our experiences and analysis useful when informing their own assessment decision making.

Introduction and Motivation

High stakes exams are a source of stress for students [1, 2], have shown to uphold systemic disadvantages for certain minority groups [3–9], and create an excessive workload for the instructional staff of a large, introductory computer programming course. During the COVID-19 remote-only teaching of the 2020-2021 academic year, the standard recommendation was to switch to more frequent, lower stakes assessments with comparatively flexible accommodations to account for the varied circumstances that students found themselves [10].

However, we, the instructors of this large introductory programming course, were wary of switching to more frequent, lower stakes assessments without considering how such a change would affect the overall student and instructional staff experience in the course. We worried that adding additional “exam-type” assignments would further penalize students who were already

penalized by the structure of traditional exams. We attempted to find research that actually analyzed whether the more frequent, lower stakes assessments were more equitable, but we were unable to find much that was applicable to our situation. Therefore, we analyzed our own course data with the intent of contributing our analysis to the existing literature.

Historic Structure of Our Large, Introductory Programming Course

Here, we provide context for the study by detailing the relevant details of the course before shifting from two exams to four low stakes assessments.

This introductory programming course of approximately 700 students/semester is typically staffed as the following: 3 faculty instructors, 9 graduate teaching assistants, 8 undergraduate teaching assistants, 1 grader, 1 administrative assistant. The assignments in the course were generally weighted as: Weekly Reviews (15%), Labs (15%), Projects (30%), Exams (40%).

The course has used weekly reviews as low stress/low-level formative individual assessments as a check to identify students who are struggling with the course material prior to taking the exams (for retroactive analysis, we define "struggling" as students who accumulate fewer than 80% of available course points). The two exams that we used (midterm and final, each worth 20% of a student's final grade) were considered to be the primary assessment of students' individual understanding of course material, as the bulk of other course work is either done in groups, with a partner, or graded leniently (e.g. "for completion").

Faculty Workload: Writing Exams Is an Excessive Drain on Faculty Time

This course is team-taught, and the faculty spend most of their time coordinating amongst each other and the teaching assistants to ensure that we provide a consistent and inclusive teaching experience across all of the various lectures and lab sections. We, the staff, also want to maximize our "student facing" time – time spent actively teaching in lecture, lab, or office hours – both because this time is the most impactful for students [11] and because this is typically the most enjoyable time for us [12, 13]. Time and effort spent on content creation for labs and lectures is also considered fulfilling as it directly supports our efforts in teaching [14].

Unfortunately, the creation of new paper-based exams each semester seemed to take up most of the faculty instructors' available "content creation" time, making it infeasible for faculty to be significantly involved with revising routine lectures and labs. Additionally, significant effort was needed by the teaching assistants to take preliminary versions of the exams to find typos and clarify instructions prior to the students taking the exam, as well as manually grade the exams after the exams were taken. Faculty felt that the creation of the exams was the most time consuming part of the exam process.

The faculty also felt that creating new exam questions each semester was significantly taking away from their available time to manage the teaching assistants and interact with students via email, office hours, or the course discussion forum. We verified this impression by systematically recording the amount of time instructional and administrative staff spent on course-related tasks during two separate weeks of the Winter 2021 term. The time logs revealed the fact that faculty were spending a total of 25 hours on Week 1 and 14 hours on Week 2 on administrative and

exam-related tasks, much more time than on student-facing activities. We thus were motivated to pursue a more sustainable assessment model that would provide the faculty with more time for student engagement.

COVID-19: An Opportunity for Change?

For several years, the faculty instructors had discussed the possibility of giving computer-based exams, rather than the comparatively time-consuming paper-based exams, but there were valid concerns around wide-spread cheating with a computer-based exam. The COVID-19 pandemic forced a shift to fully remote teaching and learning, including online computer-based exams. This was an unprecedented opportunity to consider an improvement in the overall assessment strategy for student learning in this course.

In considering alternative methods of assessing students' individual knowledge and understanding, we were often told that frequent, lower stakes assessments were preferable to infrequent, high stakes assessments. We decided that an initial step towards true "frequent, lower stakes" assessments would be to "promote" our weekly reviews into *bi-weekly* (once every two weeks) reviews. The bi-weekly reviews would serve as a more formal assessment of student understanding of the course material presented in the preceding two weeks. The bi-weekly assessments (BWAs) were multiple choice question quizzes hosted in our LMS. The BWAs were not timed; however, unlike the weekly reviews, the questions now had a small amount of randomization and students only had two attempts per instance.

Simultaneously, we revised the course's assignment weights to capture the larger amount of asynchronous weekly work the students would be doing in a fully remote semester; a comparison of the breakdown is shown in Fig. 1. The effort previously put into the low level weekly reviews was instead put towards more formal weekly asynchronous learning. The effort put into exams, the only previous method for formally assessing individual student learning, was split into the new bi-weekly assessments (15% of a student's grade) and lower-stakes exams (25% of a student's grade). The exams were computer-based, timed, with some degree of randomization.

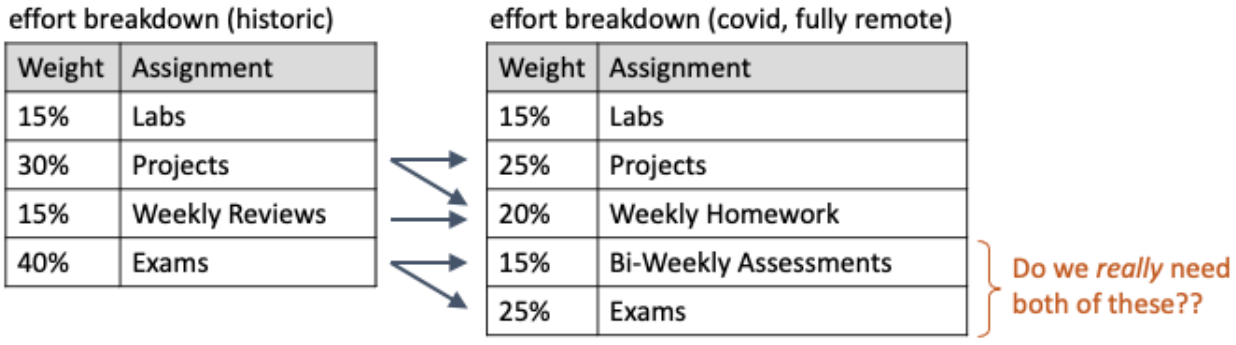


Figure 1: Comparison of student effort per assignment category, historic vs. COVID-19 / fully remote.

This course arrangement worked well for remote-only teaching and learning, but as we began to think about the return to in-person learning, we considered what our assessment strategy should

be going forward. How could we leverage the online environment to avoid a return to time-consuming paper exams, while addressing the concerns that prevented us from using them before the pandemic?

Moving Forward: An Equitable and Sustainable Assessment Strategy

The results and discussion sections of this paper will show that our analysis of past assessment scores indicates that computer-based assessment, with questions that are re-used from semester to semester, could be an effective form of assessment with a significantly reduced faculty effort once the initial question bank was created. We therefore proceeded with the change from fewer, higher-stakes, paper-based exams to more, lower-stakes, computer-based assessments.

We chose to implement a four-assessment structure for Fall 2021. The four assessments would be timed in such a way to provide both formative assessments on newly-introduced concepts and summative assessments on concepts that students had used in their projects. The assessments are spaced roughly equally apart, with the fourth assessment occurring at the end of the semester as the final summative individual assessment. The choice of four assessments was motivated partially because students reported feeling stressed about "always having to be thinking of an assessment" with the BWAs and partially because we teach two programming languages (MATLAB and C++, currently). Four assessments would give us a formative and summative assessment for each language.

The conversion of the exams and bi-weekly assessments into the four assessments allowed us to reconfigure the overall student effort breakdown to better align with the work students were doing in the course (Fig. 2).

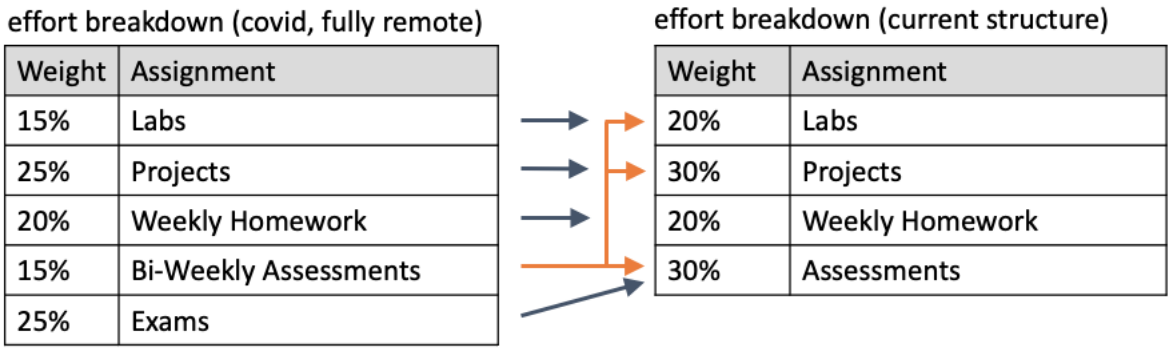


Figure 2: Comparison of student effort per assignment category, COVID-19 / fully remote vs. current course structure implemented in Fall 2021.

The current effort breakdown puts the projects back at their historic value of 30% of the course grade. The lab percentage was increased from 15% of the course grade to 20% of the course grade to reflect the greater effort required for the collaborative group lab assignments introduced in Fall 2020. For context, the weekly homework includes self-paced, asynchronous work completed via an interactive learning platform (material developed by course staff) as well as attending one lecture per week and completing a reflection form (i.e. a flipped-classroom style approach).

After evaluating many different platforms, we decided to use PrairieLearn, an online problem-driven learning system for creating homeworks and tests [15]. The Appendix includes more information on our assessment platform decision, our rationale for our choices in assessment settings, and how we describe the assessments to our students.

This assessment strategy worked well for faculty, but does it also help mediate the inequities in traditional exam strutures like we hoped? We analyzed our own data to find out.

Methods

In this section, we describe our process for analyzing data to inform decisions about revising the assessment structure, and how we assessed the effect of these changes after they were made. Our goal was to make data-informed decisions about how we enable students to demonstrate their learning while making sure our assessment practices are effective and equitable. We also wanted to utilize these data to identify ways to reallocate instructional team workload to improve the sense of meaningful work for both faculty and student instructors.

> **Research Question:** Do high stakes exams provide critical and unique information for assessing students' individual understanding of course material as compared with other forms of individual assessment?
>
> **Hypothesis:** We predict that traditional, high stakes exams (with new, high quality questions written each semester) will not provide the teaching team with critical and unique information for assessing students' individual understanding; therefore, traditional exams may be removed in favor of more frequent assessments in the style of "threshold grading" (often referred to as "mastery grading").

In this paper, we use data from Academic Years 18-19, 20-21, and the Fall Term of 2021 to explore how student exam scores evaluate student learning in the context of other assessment mechanisms. Because of the circumstantial disruption imposed by the 19-20 academic year and the switch to emergency remote-only teaching in March of 2020, we have not included those data in this analysis.

Our process for this analysis was:

- Analyze base line exam data for Academic Year 2018-2019 (a pre-COVID-19 pandemic "typical semester") for equitable outcomes across various student groups
- Analyze "intermediate assessment iteration" comparison data between higher stakes exams and lower stakes biweekly assessments for Academic Year 2020-2021 for both equitable outcomes across various student groups and equitable outcomes between exams and biweekly assessments
- Analyze "final assessment iteration" data for Fall 2021 for equitable outcomes across various student groups and for comparison to past assessment structures

The results of the base line analysis and intermediate assessment analysis informed the decisions for the structure and frequency of assessment in the final assessment iteration.

Our deep-dive into the relationship between assessments, student learning, and grade outcomes began as part of a large-scale course redesign effort, which was planned prior to the COVID-19 pandemic. However, this was not initiated until May 2020, when we were already in the midst of shifting the course to remote instruction. Our investigation of grade data focused on three central questions:

1. How equitable are our assessments?
2. How can we use available data to identify students who may go on to struggle in the course early in the term in order to implement interventions that would result in more positive outcomes?
3. What (if any) additional value was provided by high-stakes exams that could justify the amount of instructional time needed to develop them?

To attain a baseline of the grade landscape that would inform our redesign, we analyzed gradebook data from AY 18-19. We chose this as our baseline rather than AY 19-20 because we wanted to have a full academic year of data, and the sudden shift to remote learning in the middle of the WN 20 term due to COVID-19 made those data an inaccurate reflection of the baseline for a number of reasons including default masking of grades, inequitable access to technology, disruption of curriculum, and the inherent stress of an emergent pandemic.

We analyzed the exam and BWA scores for two things: correlation between the two categories and whether the exams could be eliminated in favor of something like more frequent online assessments. To simulate "removing the exam component", we set the percentage of the exam scores on the final grade to zero and set the percentage of the BWA scores on the final grade to include the percentage of both the original BWA scores and the exam scores. The final student grades were then recalculated and analyzed for equity.

Our preliminary analysis of the new assessment strategy (four assessments vs. two exams) focused on a comparison of scores across the three assessment strategies: baseline, intermediate, and final assessment iteration. We compared assignment category scores in each dataset as well as final course grades to measure the equity of our new assessment strategy.

Results

Our results of the three analyses are presented in the following sections.

Baseline Data Analysis (AY 18-19)

Across all assignment types, exams showed by far the largest outcome disparities for historically underrepresented/marginalized students across identity variables including sex, race/ethnicity, and educational generation status. These outcome disparities are summarized in Figure 3; they were calculated by subtracting the historical majority group (male, continuing generation, and non-URM[1]) from the minority group (female, first generation, and URM). A negative value

[1]We recognize that the label "underrepresented minority (URM)" is problematic, but our institution is still in the process of transitioning away from this label, and so this is the category that is available to us in our institutional data.

(pink) indicates that the minoritized group experienced lower outcomes than their peers. A positive value (green) indicates that the minoritized group experienced higher outcomes. A plethora of data point to inductive bias in high stakes exams (such as standardized exams) being responsible for such inequities [16], and thus we believe that it is possible that our course exams may also be susceptible to such implicit biases as well.

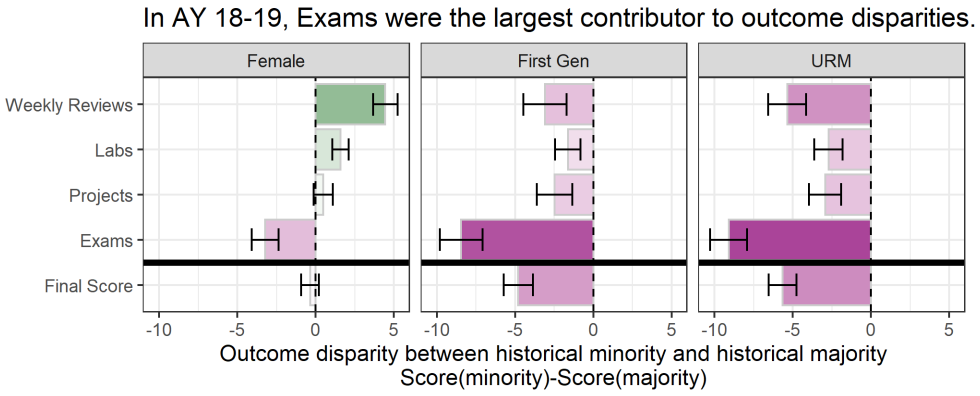


Figure 3: Outcome disparities between historically minoritized students and majority counterparts across all assignment types from AY 18-19.

Exams were the only assignment for which female students experienced lower outcomes than male students. While this difference was small (3%), it was statistically significant (Wilcoxon and T-test). First generation and URM (here defined as students with racial/ethnic identities historically underrepresented in STEM) students scored an average of 8-9% lower than their peers on exams. This disparity was particularly extreme when we disaggregated by race: the mean exam score of Black students was fifteen percentage points lower than the course mean (compared to a 5% difference on Projects, 4% on Weekly Reviews and 3% on Labs).

Intermediate Assessment Iteration Analysis (AY 20-21)

Exams remained the least equitable form of assessment for marginalized identity groups (female, URM, and first generation students), but the mean disparity was less than in Fall 2019. BWAs were not perfectly equitable, with some outcome disparities present, particularly when disaggregated by race, but they were more equitable than exams. Furthermore, the correlation between BWA and Exam scores was higher than other assessment types in these or previous semesters; the correlation coefficient was 0.61 in Fall 2020 and 0.75 in Winter 2021.

Despite this reasonably high correlation between BWAs and Exam scores, our hope to leverage the first BWA to more accurately identify students in potential need of intervention after the first two weeks was not fulfilled. We explored several numerical and logistic regression models with early-term gradebook data at the end of Fall 2020, but the small total number of struggling students (6%) led to wide variability in the prediction, and we found the models unsatisfactory mechanisms for informing targeted outreach.

Figure 4 shows the actual course score students received according to the effort breakdown table on the x-axis and the y-axis shows the hypothetical course scores if the exams were worth 0% and

the BWAs 40%.

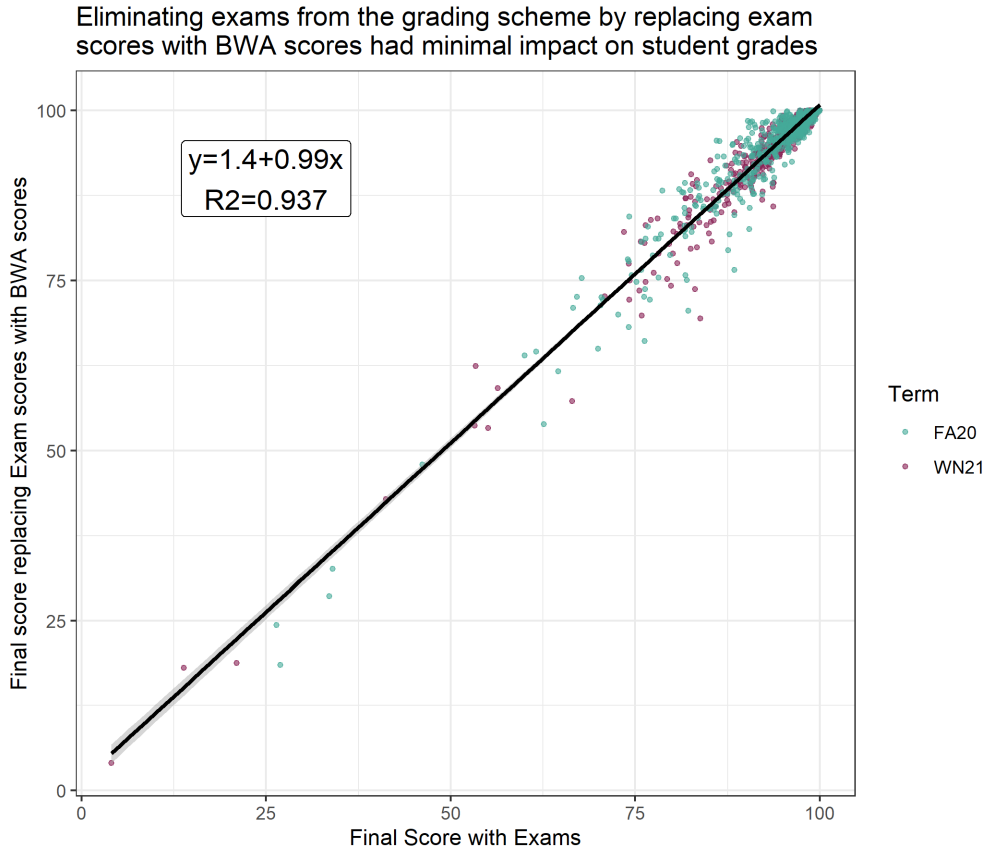


Figure 4: We evaluated whether BWAs could act as a replacement for individual learning assessment provided by high-stakes exams by hypothetically shifting all the grade weight of exams into BWAs. The result suggested that replacement of exams with more frequent lower-stakes assessments would not substantively change the existing grade distribution.

This major hypothetical shift in the grading scheme resulted in minimal changes to the individual grades students would have received when considered in context with all the other types of assignments.

Comparisons Across Assessment Strategies

Analysis of these four assessments in the Fall 2021 semester showed more equitable outcomes with respect to race/ethnicity, biological sex, and educational generation status. In our original assessment scheme, the greatest disparities emerged when we disaggregated by race/ethnicity. Figure 5 shows the mean grades across the various individual summative assessment schemes for each stage of the redesign, looking specifically at race/ethnicity (inset numbers show the number of students represented in each group). Outcomes for other identities showed similar improvement, with first generation students going from an 8 point disparity to equivalent outcomes, and female students outperforming their male peers (as they have historically done on

non-exam assignments, as shown in Figure 3).

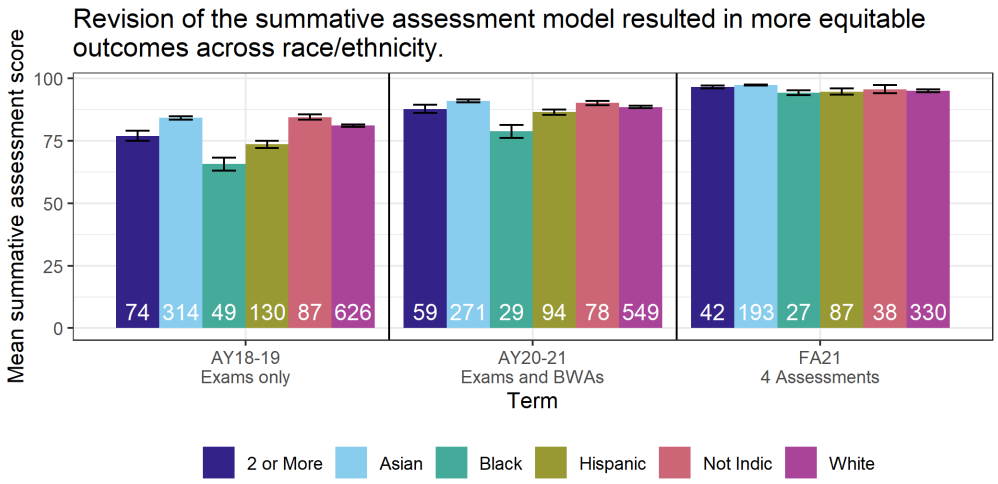


Figure 5: The evolution of assessment outcome equity with respect to race/ethnicity over the trajectory of our assessment redesign. The inset N values are the numbers of students in each group.

In addition to showing better equity outcomes, the data in Figure 5 also show that students overall received higher grades on their individual assessments in FA21 than in previous terms. This led to an increase in final scores students received in the course (summarized in Figure 6).

Figure 7 shows the outcome disparities from the earlier terms already shown in 3 in direct comparison with the most recent term.

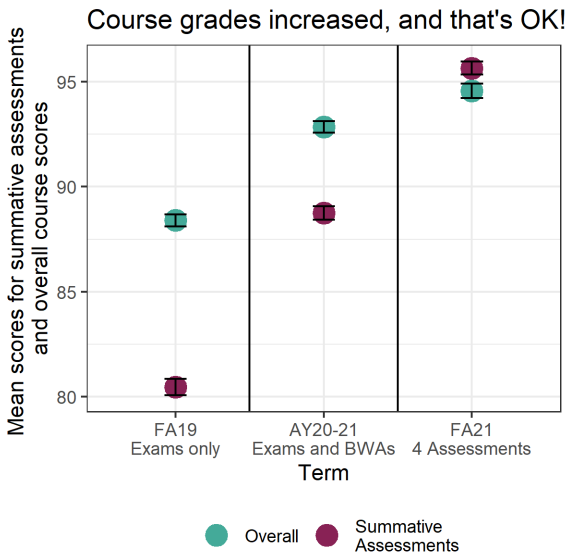


Figure 6: As students scored higher on their summative assessments, overall grades went up as well. In addition, the relative means of summative assessments and overall scores became closer in value.

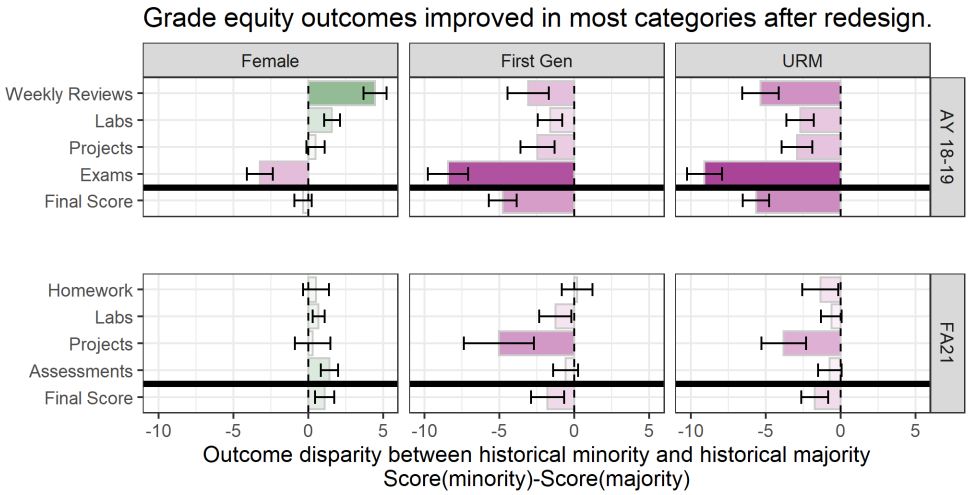


Figure 7: Outcome disparities between historically minoritized students and majority counterparts across all assignment types from AY 18-19.

Discussion

Our observations and thoughts about the equity analysis and student support capability of these assessment strategies is presented here.

More Equitable Outcomes with New Assessment Strategy

Given that exams in the traditional course structure (the base line data set) time were worth 40% of a student's grade, the disparities in the exam scores (Fig. 3) propagated to the final score, thereby perpetuating the systemic inequality experienced by many minority students. We also recognize that these disparities arose from deficiencies in the *course design* that were inequitably supporting student learning at the time.

Switching from Exams to Assessments had a particularly large impact in reducing or eliminating overall outcome discrepancies. Projects show a larger discrepancy for First Gen and URM students, despite the projects themselves remaining unchanged from the earlier terms. We hypothesize that this is more about time management than demonstration of learning, especially as students in this group are still learning during the COVID-19 pandemic and the additional stress that has added to everyone. Although students were back on campus for the Fall 2021 term, students from marginalized groups continue to experience the negative impacts cited earlier in this paper disproportionately more than their peers.

Better Early Intervention for Students Who Are Struggling

As discussed earlier, the goal of the course is to provide students with programming skills rather than stratify learners across a proscribed grade distribution.

In the base line data analysis, we found that the identities of students who struggled were not representative of the course as a whole, and exhibited a strong intersectional impact. Looking at students across both URM and First Generation (FG) Status, we found that 50% of students who were both URM and FG ultimately struggled in the course, compared to only 9% of students with neither of those identity traits. This effect was intermediate for students with a single historically marginalized identity trait (19% of non-URM first generation students and 17% of URM continuing generation students received a course score below 80%). Clearly the course's old structures for supporting student learning were working better for some students than others.

The previous assessment strategy (two exams: midterm and final) provided little information that would be useful for predicting early-on whether students would ultimately struggle in the course. Across all assignments, students who accumulated a course score of less than 80% exhibited a "slow slide", in which the gaps between students would go on to struggle and those who did not expanded gradually throughout the term. On Labs and Homework assignments during the first few weeks of the term, students who would go on to struggle averaged 1% or less below the class mean, falling to 3-4% at the end of the term. Project scores for these students started at 5% below the mean, and extended to 15% by the midterm (on which they averaged 15% lower than their peers). At this point, they were halfway through the semester and may have felt demoralized and/or hopeless about the possibility of grade recovery.

By providing students with the opportunity of demonstrating and reflecting on their knowledge earlier in the term, and allowing them to go back for additional study on specific topics before retaking the assessment (threshold grading), the new assessment model promotes earlier course correction.

In Fall 2021, 70% of the 33 students who ultimately accrued a score of less than 80% in the course scored below the mean on the first assessment, compared to 28% of the non-struggling population. Naturally, going by the class mean in a course where such a small number of students struggle is going to produce a lot of false positives by this metric. However, if we take other early-term assignments into account we can narrow the field somewhat. For example, if we flag an assignment as "low" if a student's score is a standard deviation or more below the class mean on the first three homework and projects and simply below the mean on the first assessment, we can flag only those students who receive "low" outcomes on two or more assignments. In Fall 2021, these criteria were met by 73% of students who went on to struggle, but only 8% of those who did not. We then generate a "watchlist" of students who may be having difficulty with the material; the watchlist is updated each week based on rolling criteria of the most recent assignments. The instructional staff can then contact the student with the specific details of what the student is missing and work with the student to get caught up.

Limitations

In this study, several factors were beyond our control and other factors were revealed after analyzing our results. Here, we discuss these threats to validity.

This paper reports the final outcomes from only a single semester, which was the pilot for our new assessment scheme. We are still exploring how best to use early-term data to provide early interventions for students who are not on track to receive an 80% or higher in the course. Nor do we yet have the data to analyze the longitudinal impact of this shift to see whether students are more likely to persist into the next courses in the computer science sequence and what outcomes they experience relative to the prior model. While the new assessment scheme showed much greater equity with respect to grade outcomes, particularly for students with racial and ethnic identities that are historically underrepresented and marginalized in Engineering, we must acknowledge that our institution does not have large representation among these identities, as shown by the inset N values in Figure 5. We hope to see these results confirmed and reinforced over future semesters as our data set becomes more statistically robust.

We acknowledge the possibility that our attempt to create threshold-grading-based, equitable assessments resulted in assessments that were "too easy" and so all students were able to earn high scores. It is possible that the more equitable scores shown in Fig. 5 are an artifact of assessment questions that perhaps do not give students a true assessment of their deeper understanding of course material. However, we do note that the new assessment strategy includes the ability for students to retake the assessment if they receive $< 90\%$. As we encourage students to take advantage of this policy, we would expect that the majority of students would earn a good final score on the assessment.

This course may be anomalous in its role as a "weed in" course. The goal of the course is to introduce students to computing in engineering so that they can succeed in their later courses; it is

not meant to find only the "best" programmers so that they may be funneled into the Computer Science and Engineering department. While the course is required for all first-year students in engineering at our institution, it is generally perceived favorably by most students and is not considered to be a hard course for the majority of students. In the last 5 years, 74% of students earned an A- or higher in the course, and our goal for the course is that students learn. As such, we feel that we have the privilege to prioritize skill-based assessments and threshold-grading with many opportunities to develop competency in course material. If everyone gets an A because they can do all the things we want them to do, then that means we have done our jobs well!

Lastly, we must also recognize that with the exception of our baseline data from AY 2018-2019, the semesters from which these data are drawn took place during the COVID-19 pandemic, an unprecedented time in higher education. For AY 20-21, students and instructors were forced into a remote learning environment in which many students experienced additional stress, anxieties, and other responsibilities such as family care or shifting financial responsibilities. These challenges as well as access to reliable internet and productive study spaces were not experienced equally by all students [17]. Furthermore, the later years in which the data were collected took place amid not only a global pandemic, but a time of political and social upheaval in the United States. The learning context is therefore (hopefully) irreproducible, but we hope that the positive outcomes we have observed will persist into less tumultuous times.

Conclusions and Future Work

This study examined exams and alternate assessment strategies for a large, introductory computer programming course for Academic Years 2018-2019 and 2020-2021. Our goal was to examine our assessment mechanisms to make an informed choice as to the best overall assessment scheme for the course beginning in Fall 2021.

Conclusions

Our analysis of the baseline AY 2018-2019 data revealed that students who receive fewer than 80% of available course points exhibit performance gaps across almost all assessment mechanisms, implying that exams are not providing additional data to identify students who are not absorbing the material. Additionally, the exams were found to perpetuate structural racism and sexism through inequity in exam outcomes.

Our analysis of the AY 2020-2021 data revealed that low-stakes bi-weekly assessments provide similar learning data to exams. An investigation wherein we simulated eliminating the exam component of the overall grade by assigning that percentage of the students grades to the bi-weekly assessment scores increases mean and median scores by 2-3%, but has minimal impact on whether students are defined as "struggling."

Given the workload of exam preparation, cheat-checking, etc., exams are taking up a disproportionate amount of instructional team time without substantively contributing to the discrimination between higher-learning students and lower-learning students.

Therefore, we do *not* reject our hypothesis. We instead elected to create high quality, but automatically randomizable assessments that focus on the core skills we hope to teach our

students. More frequent assessments allow us to intervene sooner with students who are starting to struggle in the course. The structure of the randomization of the assessment questions, along with the policy of not releasing assessment answers to students, allows the questions to be re-used in future semesters, thus removing the burden of exam creation from the instructional staff and allowing them to spend more time with their students.

The preliminary analysis of the new assessment structure indicates improved grade equity in assessments and a better match between assessment scores and overall course grade.

Future Work

This study has given us several new directions to investigate as we continue to work towards a sustainable, equitable, and inclusive course.

The concerns about the assessments being "too easy" are legitimate. We plan to do a further analysis of assessment scores that includes looking at the distribution of student scores on the original assessment and then including the effect of students retaking the assessment. A study that "follows" our students after taking this course to see their performance in subsequent computing classes might also reveal whether "too easy" assessments harm the students later on.

We also intend to carefully plan the increased level of randomization of the questions on the assessments that we are now using in the course. Randomization is an important tool in deterring cheating; however, too many "degrees of freedom" in selecting the questions lessens the effectiveness of the practice assessments and can increase student stress while taking the actual assessment. At the same time, the ability to retake an assessment to reach a threshold score is a powerful lever for students to pull, both to lessen stress on assessments and to allow them to actually assess their own competency in course material.

We hope that it is possible to combine everything ("harder" questions that require a deeper understanding of course material, more questions to pick from in the question bank, and allow retakes) to create a truly effective assessment mechanism that is both sustainable for instructors to maintain and gives all students confidence in their competency in course material.

Using more frequent formal assessments gives us the opportunity to improve our student support. Our next step in this aspect of our work will be to continue to refine our identification strategies for students that are struggling, and then work with collaborators in Behavioral Science to implement nudges or outreach messages that are in line with educational best practices.

Acknowledgement

This research was supported by the Foundational Course Initiative at the University of Michigan under IRB HUM00150716.

References

[1] Reda Abouserie. Sources and levels of stress in relation to locus of control and self esteem in university students. *Educational Psychology*, 14(3):323–330, 1994. doi: 10.1080/0144341940140306.

[2] David Robotham and Claire Julian. Stress and the higher education student: A critical review of the literature. *Journal of Further and Higher Education*, 30(2):107–117, 2006. doi: 10.1080/03098770600617513.

[3] Noe Medina and D.Monty Neil. Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in America's public schools. third edition (revised). *Educational Resources Information Center (ERIC)*, pages 139–153, March 1990.

[4] Carolyn A. McBride Davis, John R. Slate, George W. Moore, and Wally Barnes. Advanced placement exams, incentive programs, and cost effectiveness: A lack of equity and excellence for black students in texas, new york, and florida. *Journal of Negro Education*, 84(2):139–153, Spring 2015.

[5] Drew S. Jacoby-Senghor, Stacey Sinclair, and J. Nicole Shelton. A lesson in bias: The relationship between implicit racial bias and performance in pedagogical contexts. *Journal of Experimental Social Psychology*, 63: 50–55, 2016. doi: 10.1016/j.jesp.2015.10.010.

[6] Ilja Cornelisz, Martijn Meeter, and Chris van Klaveren. Educational equity and teacher discretion effects in high stake exams. *Economics of Education Review*, 73:101908, 2019. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2019.07.002. URL https://www.sciencedirect.com/science/article/pii/S0272775718306976.

[7] Jorge A Girotti, Julie A Chanatry, Daniel M Clinchot, Stephanie C McClure, Aubrie Swan Sein, Ian W Walker, and Cynthia A Searcy. Investigating group differences in examinees' preparation for and performance on the new mcat exam. *Academic Medicine*, 95(3):365–374, 2020.

[8] Catherine Reinis Lucey and Aaron Saguil. The consequences of structural racism on mcat scores and medical school admissions: the past is prologue. *Academic Medicine*, 95(3):351–356, 2020.

[9] David M. Quinn. Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3):375–392, 2020. doi: 10.3102/0162373720932188.

[10] College of Engineering: Minimum quality standards for remote exams and assessments. *NEXUS, College of Engineering, University of Michigan*, Apr 2020. URL https://docs.google.com/document/d/1gyVssMuu8imY8N5j_rtyEeEgpFv4QdU7hjIkZObLeow/edit#heading=h.gjdgxs.

[11] Peter A. Daempfle. An analysis of the high attrition rates among first year college science, math, and engineering majors. *Journal of College Student Retention: Research, Theory & Practice*, 5(1):37–52, 2003. doi: 10.2190/dwqt-tya4-t20w-rcwh.

[12] Jantine L Spilt, Helma MY Koomen, and Jochem T Thijs. Teacher wellbeing: The importance of teacher–student relationships. *Educational psychology review*, 23(4):457–477, 2011.

[13] Anne C. Frenzel, Betty Becker-Kurz, Reinhard Pekrun, Thomas Goetz, and Oliver Lüdtke. Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology*, 110(5):628 – 639, 2018. URL http://proxy.lib.umich.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2017-57177-001&site=ehost-live&scope=site.

[14] Kent A. Crick, Elise A. Frickey, Lisa M. Larson, and Mack Shelley. The role of teaching self-efficacy in electrical and computer engineering faculty teaching satisfaction. In *2020 ASEE Virtual Annual Conference Content Access*, Virtual Online, June 2020. ASEE Conferences. URL https://peer.asee.org/35366.

[15] Matthew West, Geoffrey L. Herman, and Craig Zilles. PrairieLearn: Mastery-based online problem solving with adaptive scoring and recommendations driven by machine learning. In *2015 ASEE Annual Conference & Exposition*, Seattle, Washington, June 2015. ASEE Conferences.

[16] Noe Medina and D Monty Neill. Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in america's public schools. (revised). 1990.

[17] United States Department of Education. Education in a pandemic: The disparate impacts of COVID-19 on America's students, June 2021. URL https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.pdf.

[18] PrairieLearn. PrairieLearn FAQ: Remote Exams, accessed Feb 13, 2022. URL https://prairielearn.readthedocs.io/en/latest/remoteExams/f.

[19] Christopher D. Schmitz, Geoffrey L. Herman, and Timothy Bretl. The effects of second-chance testing on learning outcomes in a first-year stem course in engineering. In *2020 ASEE Virtual Annual Conference Content Access*, Virtual Online, June 2020. ASEE Conferences.

[20] Ronald F DeMara, Shadi Sheikhfaal, Paul J Wilder, Baiyun Chen, and Richard Hartshorne. Blueshift: Rebalancing engineering engagement, integrity, and learning outcomes across an electronically-enabled remediation hierarchy. *Computers in Education Journal*, 10(1):1–12, 2019.

Appendix

This Appendix provides further details on our choices and rationale for assessments and assessment settings for this large, introductory computer programming course.

Assessment Platform Requirements

The form of the assessments and the choice of delivery platform are crucial to the long-term success and sustainability of the assessments in this course. Our primary requirements (in no particular order) for the assessment platform were:

- Supports a variety of question types, including "mixed up code" or "select the lines of code to make a program work" type questions
- Support for code font and syntax highlighting
- Ability to create different "versions" of a question by using a template question and placeholders for different versions of variable names, figures, etc.
- Ability to create randomized questions for individual students
- Easy to maintain a repository of questions, with subdirectories and version control
- Ability to serve the assessment, grade the assessment, and return the results to the student without interaction by the teaching staff other than the initial set up
- Ability to easily grant deadline extensions
- Ability to scale up to 1000 students with no issues (within reason)
- Availability of prompt and professional help, especially as we adopt a new platform
- Ability to re-use questions from semester to semester without egregious or widespread cheating

Structure and Logistics of the New Assessments

A common concern with computer-based exams and assessments is the threat of cheating. PrairieLearn provides guidance on the pros and cons of various choices of assessment settings [18]. PrairieLearn's recommendation is to use synchronous, timed assessments for exams. However, there were several barriers to our adopting this format:

- We do not have a computer-based testing facility at our disposal, therefore there is no secure location where all 700 of our students could take the assessment at the same time.
- It is difficult to reserve enough classrooms across campus at the same time, so it is hard to even find enough space for all our students to be in at the same time.
- Even if we could find enough classrooms at the same time, electrical outlets are limited, and we did not want students to have to worry about power for their computers while taking the exam.
- We knew from student feedback during AY 20-21 (the year of fully-remote teaching due to COVID-19) that finding a quiet place with reliable internet access to take an exam at a specific time was very difficult for some students, therefore we did not want to add that stress burden to students taking our assessments.

Due to these limitations, we ultimately chose to use asynchronous, timed assessments for our new assessments. If a computer-based testing facility is built at our institution, then we will revisit this choice.

Description of Assessments Given in Syllabus

[This is how we describe the assessments in our syllabus. More information on the logistics of taking the assessments is released separately to students when the practice assessments open.]

There will be four assessments in [COURSE]. Each assessment will cover the topics used so far in the course. If the topics have only been covered by [ONLINE ASYNCH PLATFORM] work (and not projects yet), then the assessment will cover those topics in a little bit lighter way; we can think of this as a "Level 1" understanding of the topics. If the topics have been covered in a project that is due before the assessment, then the assessment will cover those topics in the same detail as the projects; we can think of this as a "Level 2" understanding of the topics. Here's a table that compares the two levels:

Assessments will be hosted on the online platform PrairieLearn. More information about assessments and PrairieLearn will be provided separately from this syllabus. Key things to know about the assessments:

- There will be practice assessments that you can take multiple times; the practice assessments will help you see what we mean by "Level 1" and "Level 2" questions (we're finding that it's hard to explain what we mean by those levels in words, but easier to show people examples of questions!).

	Level 1 Questions	**Level 2 Questions**
When in the learning cycle?	This type of assessment question enables you to demonstrate that you have learned a particular skill or concept before you are asked to use it in the context of a project.	This type of assessment question enables you to demonstrate how to apply skills and knowledge in more complex ways after you have used them in the context of a project.
What are they designed to assess?	These questions will help you gauge how well you can demonstrate individual knowledge and skills you will need in the course, and give you a chance to identify topics that may require additional practice.	These questions let you show deeper levels of understanding and procedural knowledge by strategically combining multiple concepts and/or skills to solve real-world coding problems.

- The assessments themselves will be open note/open computer (you just can't talk to anyone else or otherwise collaborate with anyone on the assessment).
- You will be able to immediately see your score for the assessment after you finish the assessment, but you won't be able to see your answers; you can come into office hours, though, and we're happy to go through your assessment answers with you.
- If you earn $< 90\%$ on an assessment, you can come in to office hours and review your answers with a staff member so that we can help straighten you out on whatever concepts you got wrong; after this meeting, you will be able to retake the assessment and earn up to the threshold score of 90%.

Assessment Design

The following sections provide detail and rationale for our choices in assessment settings.

Assessment "Open Window"

The assessments are open for 3-4 days (depending on what the academic calendar is), and students may start the assessment at whatever time they wish during the open window. This large open window has been helpful for students as they can arrange their assessments for this course around the other assessments and exams they have in other courses, so we do not need to offer an alternate time for the assessments.

Assessment Time Limit

The original exams were each two-hour exams. Therefore, we elected to keep the same amount of total time the same: 4 hours. Each of the four assessments is designed to be comfortably completed by a well-prepared student in 60 minutes or less. We provide 75 minutes to allow for

technical issues and to reassure students that there is no need to rush through the assessment.

Practice Assessments

Each assessment has a corresponding practice assessment that is very similar to the actual assessment, including having a time limit; the practice assessments can be taken as many times as the student wishes and students can see their graded answers. The practice assessments serve multiple purposes:

- They allow students to become familiar with the PrairieLearn platform prior to taking an actual assessment.
- They allow students to learn the types of questions we ask on assessments, how the questions are organized, where the reference material is shown, etc.
- They allow students to see how the randomization in the questions work, so that they know to read new questions carefully and not make assumptions based on a previous version of a question.
- They allow students to give themselves a "knowledge check" prior to taking the actual assessment.
- They allow students to essentially give themselves "unlimited retakes" on the assessment in case the first practice assessment shows a deficiency in an area of learning.
- They allow students to practice and improve their time management skills, since the practice assessment has the same time limit as the actual assessment (including extended time accommodations).

We chose to allow the practice assessment to be available during the actual assessments. This decision was primarily motivated by what we consider to be programming best practices: one of the skills we are trying to teach students is how to apply "common patterns" to solve an engineering problem using computing. We realized that having the practice assessment open and using it for pattern recognition is no different than having our examples from the asynchronous weekly homework or labs ready as reference material. As our goal is for students to learn and practice this skill, the practice assessments themselves become approved reference material for the assessments.

Assessment Results

The assessments are set to be automatically graded "all at once" after the student has submitted their assessment. Because the assessments are timed, we wanted students to complete the entire assessment before seeing their score. If they used the "grade as you go" option, and they got a question wrong, they may feel disheartened and do worse on the rest of the assessment. Once a student submits their assessment, they immediately see their final score but, unlike the practice assessments, *students do not see their answers*. This decision was motivated by the need to re-use assessment questions in future semesters. It is harder to upload an answer key to Chegg or CourseHero if you do not possess an answer key in the first place. If students would like to see

their answers, they can come to office hours for an assessment consultation (described below under Assessment Retakes).

Assessment Retakes and Assessment Consultations

We decided to offer a formal policy of retaking the assessment if you got less than 90% to reinforce our commitment to student learning and to allow students the chance to deepen their learning first and then re-assessing their competency [19]. If a student earns less than 90% on an assessment, they come to office hours for an "assessment consultation". During the assessment consultation, a teaching assistant confidentially meets with the student and goes through their assessment answers with them. This gives our teaching staff an opportunity to intervene with students and help them straighten out any misunderstandings [20]. After the consultation, the student is given access to the assessment retake, which is the same assessment as the actual assessment but with the score capped at 90%. All students are welcome to come for an assessment consultation to go over their answers, but anyone who earned 90% or greater is not eligible for a retake.

Missed Assessments

If a student missed taking the assessment entirely (because they forgot, not because they have an excused absence), we allow them access to the assessment with the score capped at 60%; the student is not eligible for a retake in this situation.

Concerns About Cheating

As previously mentioned, there are always concerns about cheating on computer-based assessments and exams. We have chosen to approach cheating concerns by attempting to address the most common *reasons for cheating*, rather than attempt to make an assessment that is cheat-proof (as this is impossible). Our assessment settings and choices described above all reflect this approach. We are transparent about this approach with our students. This is the statement for "Common Student Concerns About Cheating" that we include in our assessment logistics post (see the Appendix for the full post):

- [COURSE] is not graded on a curve. This reduces the temptation to cheat "because everyone else is going to cheat so I have to as well otherwise I will get a bad grade."
- Open notes/resources reduces, or even eliminates entirely, the need to memorize course content; therefore, there is less temptation to cheat by having unauthorized resources.
- This is a timed assessment and the questions are randomized. We expect this combination to make cheating by collaboration much more difficult.
- This assessment is designed to be finished with time to spare by a student who is well-prepared for the assessment. If you are organized, have reviewed your project code, and can quickly locate examples to use as a "template" for your code writing/code assembling (see the PrairieLearn practice assessment), then you should have plenty of time to complete each question and have time to review your answers.

- You can save your answers after each question. Therefore, if your internet temporarily goes out, you can return to the assessment with all of your answers saved. No need to cheat because all your answers were lost!
- The assessment may also include other unspecified countermeasures to make cheating more difficult, or at least incredibly confusing for the would-be cheaters. We wish we could share the details, because they're entertaining, but that would defeat the point.

We acknowledge that this approach may not be feasible for other courses, but it has been effective thus far for this course.

Sample of Announcements Describing the Logistics of an Assessment

We include here examples of the announcements we send out for our Assessment 1. You can see that they are highly detailed, and that transparency is important for both general inclusivity in teaching and to help fend off emails from confused students! Anyone is welcome to adapt our language here for their own uses.

Logistics Post on Piazza (the course Q&A platform)

This is an example of the logistics post we pin to the top of Piazza, the course Q&A forum. This post is added roughly one week prior to the opening of the actual assessment.

Assessment 1 Logistics

Assessment 1 is coming up this week. Please read this announcement carefully so you understand how to take the assessment.

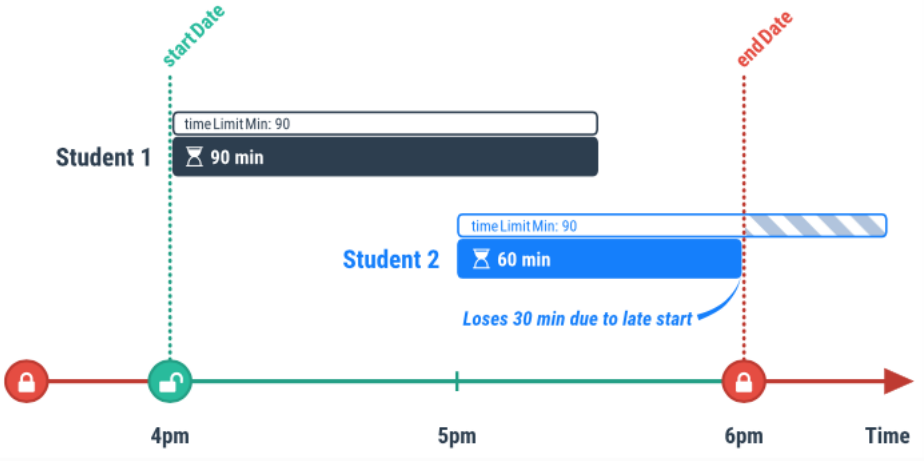
Assessment Goals
Our goal for assessments in ENGR 101 is for *you* to be able to assess your knowledge level of the course material. Therefore, we are creating assessments that directly mirror the work we have been doing in ENGR 101.

We will not create purposefully difficult assessments just so we can get an "exam distribution" of grades. We will try very hard to create a *good* assessment that thoroughly checks your understanding of the course content, especially those topics and skills that are most important for your future engineering careers.

We intend for this assessment to be a good "in the moment" indicator of your ENGR 101 knowledge and skills. We hope that you will take the assessment seriously in that you will prepare for the assessment and put forth good effort in taking the assessment. But we do *not* recommend that you spend hours and hours preparing for the assessment – that defeats the purpose of having this assessment check your "in the moment" knowledge. Some preparation and organization would be great (see below for some tips), but it is not our goal to have you overly stress about assessments in this course.

Assessment Procedure

- You will take the assessment online using a platform called PrairieLearn. Click here to access our course on PrairieLearn. If you go there now, you'll see the practice assessment. The real assessment will be available on Wednesday.
- **The assessment will be open from Wednesday, February 2nd at 12:01am until Saturday, February 5th at 11:59pm. All times are Ann Arbor times.**
- **You may start the assessment anytime during the "open" window.**
- The assessment will be randomized. Each student will receive a randomly selected version for each question, selected from a large question bank.
- **Once you start the assessment, you have 75 minutes to complete it.** The PrairieLearn system will keep time for you. (If you have extended time accommodations through the SSD office, your allowed time will be adjusted accordingly.) We have designed the assessment to be comfortably completed in less than 60 minutes, if you are a well-prepared student. We added 15 minutes extra to account for technology issues and to help make sure no one is unduly rushed in taking the assessment.
- **But it is up to you to make sure you have enough time to finish the assessment.** Here is a graphic from PrairieLearn with more detail about starting the assessment and how that may impact the amount of time you have available to complete the assessment (note that this example is for a 90 minute assessment, and our assessment is 75 minutes):



- You may answer the questions in any order.
- In line with the CoE honor code, the assessment will not be proctored (i.e. you don't need to take the assessment with a webcam).
- You will sign an abbreviated version of the CoE honor code at the beginning of the assessment, just prior to starting the assessment on PrairieLearn.
- Make sure to save your answers by clicking the "Save" button for each question.

- When you are done, you can click "Finish assessment" to close our your assessment session. If you forget to do this, PrairieLearn will automatically close your assessment when the timer runs out.

Assessment Content and Online Practice Assessment

- This assessment will cover content from Runestone Chapters 1-7, Labs 1-3, and Projects 1 & 2.
- There is a practice assessment available on PrairieLearn. Please make sure to take this practice assessment prior to taking the actual assessment so that you know what kinds of questions to expect on the actual assessment.
- You are able to "grade as you go" for the practice assessment. Please do this to help yourself understand which topics you understand well and which topics you should review prior to taking the actual assessment.
- The practice assessment IS available during the actual assessment, but please PLEASE be careful to take the actual assessment before it closes!

Asking Questions During the Assessment Window

- We have thoroughly tested all of the assessment questions, and we are confident that the questions are working the way they are supposed to. Please make sure you understand the practice assessment questions prior to taking the actual assessment, as this will help you to be well-prepared for the assessment.
- We will not answer questions about the actual assessment until after the assessment has closed, to be fair to all students taking the assessment. This includes no assessment questions during in-person office hours, zoom office hours, or on Piazza.
- We are happy to talk about the practice assessment, however, as much as you want.

Allowed Resources & Cheating

- The assessment will be "open everything" including notes, Runestone, your lab and project code, other course resources, or Internet resources, including online MATLAB documentation. You're also welcome to use MATLAB during the assessment if you find it helpful.
- You are welcome to search for anything online, but you are not allowed to solicit help from others (e.g. posting an assessment question on StackOverflow, Chegg, Slader, Course Hero, etc. asking for help).
- As always, collaboration with others in any form is prohibited and considered an honor code violation.

Common Student Concerns About Cheating

- ENGR 101 is not graded on a curve. This reduces the temptation to cheat "because everyone else is going to cheat so I have to as well otherwise I will get a bad grade."
- Open notes/resources reduces, or even eliminates entirely, the need to memorize course content; therefore, there is less temptation to cheat by having unauthorized resources.
- This is a timed assessment and the questions are randomized. We expect this combination to make cheating by collaboration much more difficult.
- This assessment is designed to be finished with time to spare by a student who is well-prepared for the assessment. If you are organized, have reviewed your project code, and can quickly locate examples to use as a "template" for your code writing/code assembling (see the PrairieLearn practice assessment), then you should have plenty of time to complete each question and have time to review your answers.
- You can save your answers after each question. Therefore, if your internet temporarily goes out, you can return to the assessment with all of your answers saved. No need to cheat because all your answers were lost!
- The assessment may also include other unspecified countermeasures to make cheating more difficult, or at least incredibly confusing for the would-be cheaters. We wish we could share the details, because they're entertaining, but that would defeat the point.

Preparing for the Assessment

- We recommend that you prepare for the assessment by getting organized. You have many resources that you are allowed to use during the assessment, but if you are not organized, you will waste time looking for information and you run the risk of running out of time.
- You can get organized by creating some quick reference guides for yourself:
  - **Common patterns** we've used in Runestone, lab, and projects. List the pattern, what it does, and at least one example that you've used.
  - **MATLAB functions.** List the function, what it does, and at least one example that you've used.
  - **Examples from Runestone and lab.** List the example how you remember it (e.g. "Finding the minimum value of a vector" or "the one with the people using scooters"). Explain to yourself in your words what this example does and then include the code for the example.
  - **Project code.** List what you've done in the projects, describe it in your own words, and make sure you can find your project code quickly for examples.

- The assessment will follow the same format as the practice assessment, so make sure you have looked at that practice assessment at least one day prior to when you plan to take the assessment.
- **Look at your schedule for this week NOW** and find a time when you will be able to take the assessment in an environment that is comfortable for you. Then, go ahead and add "take ENGR 101 Assessment 1" to your schedule so that you will remember to take it!
- Some of the labs had optional exercises at the end. If your group didn't get to those optional exercises, you can do them now as preparation for the assessment.

Assessment Retake
As previously stated, our goal with assessments in ENGR 101 is for you to check whether you have a solid understanding of the course material. It follows, then, that if you do not have a solid understanding, we want to help get you back on track! If you earn ¡ 90% on the assessment, you may retake that assessment for up to 90% of the original points. Here are the steps to retaking an assessment:

- Come to in-person office hours after the original assessment has closed
- Ask for an assessment consultation
- A GSI/IA will privately and confidentially go over your assessment answers with you and help you straighten out any misunderstandings and answer any questions that you have about the assessment
- After your consultation, you will receive access to the assessment retake on PrairieLearn
- Complete the retake before its close date (generally one week after the original assessment's close date)
- Your final score for the assessment will be whichever is higher: your original score or your retake score

To answer some anticipated questions about retakes:

- If you earn $\geq$ 90% on the assessment, you are not eligible for a retake. You are still welcome to come to office hours to have a staff member go through your assessment with you.
- Assessment consultations will only be done at in-person office hours. Consultations will not be done via Zoom, not even "just to see my answers".

We know that this is a LOT of information. Please post your questions on the assessment logistics as a follow-up discussion to this post so we can make sure everyone is as clear and comfortable as you can be for the assessment this week. You can post questions about the content of the practice assessment to Piazza as well.

Canvas Announcement for the Practice Assessment

This is an example of the announcement we send out when the practice assessment is available. The Piazza post is made first (so it can be linked in this announcement), then this announcement is sent out.

Subject: Assessment 1 (Practice) is available on PrairieLearn

Hello Everyone,

Assessment 1's **practice** assessment is available on PrairieLearn right now. You can take this practice assessment as many times as you want – your scores on the practice assessment are not reflected in your course grade in any way.

Please read through this Guide to Taking Assessments on PrairieLearn [hyperlinked in actual announcement]. The guide will likely answer a lot of questions that you will have about how PrairieLearn works and how the practice assessments differ from the actual assessments.

IMPORTANT NOTE – This practice assessment is very very similar to the "real" assessment. Assessment 1 covers everything we have done in Runestone, Lab, and Projects as of when the assessment opens. Therefore, there are some parts of the practice assessment that will likely make more sense after you have completed Lab 3, Runestone Chapters 6 & 7, and Project 2.

If you have extended time, please check that *Assessment 1 (Practice)* is giving you the correct amount of time. We have designed the assessment such that a well-prepared student can comfortably finish the assessment in less than 60 minutes; we have set the time limit for 75 minutes to account for any drops in internet connectivity, your computer just being slow, making sure you don't feel rushed taking the assessment, etc. If you have submitted a form for time and half, you should see that your allotted time is 113 minutes. If you have submitted a form for double time, you should see that your allotted time is 150 minutes. **If you do not see your time accommodations please email [assessments email address] ASAP** so we can figure out what went wrong before Assessment 1 opens.

If you have a VISA form that includes exam accommodations, please make sure you have it submitted to the Accommodate system by Monday, January 31st.

If you have questions about the PrairieLearn platform or how we're doing the practice assessment, please ask them as a follow up discussion to this Piazza post for Assessment 1 (Practice) Q&A. [hyperlinked in actual announcement]

We have also posted detailed logistics for Assessment 1 on Piazza. [hyperlinked in actual announcement]

Note on the Timing of Assessment 1
In case you are wondering why we are having an assessment immediately after Project 2 is due… well, that is on purpose, actually. Projects 1 & 2 focus on core MATLAB skills like vectors/matrices, functions, logical indexing, driver programs, and image manipulation. We

want you to be able to assess your skills on these concepts right after you've been using them in Project 2, when those skills are all fresh in your brain.

Similarly, we want you to be able to assess your *initial* understanding of the core skills that Project 3 will use: plotting, statistics, simulations, etc. If you have a solid initial understanding of these concepts, then you will have a smoother programming experience with Project 3. If you have any misunderstandings, those misunderstandings will show up in Assessment 1, and we'll be able to get you straightened out before you get stuck trying to understand Project 3. We hope this explanation helps!

The ENGR 101 Staff

Canvas Announcement for the Actual Assessment

This is an example of the announcement we send out when the actual assessment is available. We set this announcement to "delay post", i.e. we write it ahead of time and then set the announcement to go out to students when the assessment first opens.

Subject: Assessment 1 is Open on PrairieLearn

Assessment 1 is now available on PrairieLearn.

- **Assessment 1 closes at 11:59pm on Saturday, February 5th, Ann Arbor time.** Remember that it is up to you to make sure you start the assessment such that you have your full allotted time before the assessment closes. To put this another way, if you do not start your assessment by 10:44 pm EST on Saturday (2/5), you will not have the full 75 minutes to take your assessment. Please start earlier if you have extended time.
- Once you start the assessment, **you will have 75 minutes to complete the assessment** (if you have accommodations for extra time, those will be reflected on PrairieLearn).

Taking the Assessment

Assessment 1 is an individual assignment – collaboration with others or asking for help on the Internet is strictly against the Honor Code.

Please refer to the full set of instructions and guidelines for Assessment 1 on Piazza [hyperlinked in actual announcement] for details. **Make sure you read the follow up discussions on both that post and the Assessment 1 (Practice) Q&A because there may be important clarifications that have been made about the questions.**

Because this assessment is open for several days and you can take it whenever you want, we are not able to answer questions about the assessment as you take it. If you have a concern about a specific question on the assessment, you can make a **PRIVATE** Piazza post, and we will get back with you after the assessment closes.

We have done our best to make the questions on the assessment as clear and concise as pos-

sible. But we are only human and we might have messed something up. Please do your best if anything seems unclear, and know that we will manually correct grading for any questions/question parts that we inadvertently made confusing.

Recommended Procedures For a Less Frustrating Assessment Experience

- Get your resources organized before you start.
- Get in a place where you will be relatively free of distractions. If you have headphones, you could potentially use those to play non-distracting music (here is Dr. Alford's list of songs "for doing work" [hyperlinked in actual announcements]) or just block noise. Consider putting up a sign saying "TAKING EXAM" so people know to leave you alone. (This isn't really an "exam" but "exam" has fewer letters to write than "assessment".)
- Make sure you have a full 75 minute window in which to take the assessment (or more if you have extended time accommodations).
- Use a relatively updated web browser.
- Make sure to save your answers as you go along.
- Double check that you have saved answers to all your questions before you click "Finish assessment".

After the Assessment After you click "Finish assessment", you will be able to see your score for Assessment 1. If you would like to see your answers/point breakdown, please come to office hours after the assessment has closed on Saturday (see the Piazza logistics post for more details). Grades will be pushed to Canvas when we update grades after the assessment has closed.

If, after you receive your score, you are interested in a retake, please refer to the Piazza logistics post for more information on assessment retakes.

The ENGR 101 Staff