## ASEE 2022 ANNUAL CONFERENCE Excellence Through Diversity MINNEAPOLIS, MINNESOTA, JUNE 26<sup>TH</sup>-29<sup>TH</sup>, 2022 SASEE

Paper ID #37674

# Decreasing Student Stress Through Multi-Attempt Digital Engineering Assessments with Rotating Questions

## **Duncan Davis (Associate Teaching Professor)**

Duncan Davis is an Associate Teaching Professor in First Year Engineering. His research focuses on using gamification to convey course content in first year classes. He is particularly interested in using the construction of Escape Rooms to teach Engineering Principles.

## **Ciana Winston**

Fifth-year Environmental Engineering Student at Northeastern University, Boston with interests in soil remediation and green energy solutions.

© American Society for Engineering Education, 2022 Powered by www.slayte.com

## Decreasing Student Stress Through Multi-Attempt Digital Engineering Assessments with Rotating Questions

#### Abstract:

This paper will discuss building multi-attempt quizzes and exams that use the Canvas Learning Management System (LMS) to deliver engineering assessments designed to lower overall student stress and anxiety. These assessments use practice-focused questions that force students to build programs (C++ and Matlab), draft engineering drawings (AutoCAD and Solidworks), and apply engineering design, ethics, and intellectual property concepts to solve open-response questions. Each time a student takes the assessment, the questions shift randomly within topical areas facilitated by question banks that rotate among questions of similar difficulty for each new attempt. Each assessment is composed of 1-5 question banks to ensure all topics were covered throughout the various assessments. In total, 17 assessments used this framework across two fall semesters and one spring semester in the 2020 and 2021 school years.

Over the course of 10 unique assessments and 3 semesters, 54.7% of students used more than one attempt on each quiz or exam when averaged across all assessments in the study. Using LMS analytics and open-ended questions administered through an end-of-term student survey, 86.2% of students reported positive experiences regarding the assessment methods, 51.2% of students reported decreased anxiety, and 22.6% of students appreciated the greater flexibility provided by the assessments. From this data, multi-attempt assessments had a positive impact on students' wellbeing in three semesters that were particularly challenging due to the COVID-19 pandemic.

#### Introduction:

Alongside the rise of technology, digital learning management systems (LMS) transformed education at the collegiate level, further accelerated due the onslaught of COVID-19 [1], [2]. LMSs can be applied across academic boundaries and in an interdisciplinary manner due to platforms being highly customizable, accessible, and adaptive [2], [3]. Some popular platforms used within the field of engineering education include Canvas, Blackboard, WISEflow, and Moodle. Most LMSs offer space for informational content, homework, and assessments, making it an all-in-one platform that helps simplify assessment distribution and evaluation by collecting digital inputs from users [4], [5]. The statistics provided by an LMS are useful when analyzing a specific instructor-student-content relationship, making it easier to customize the interactive content and optimize the experience of all users. The customization process looks unique for each administrative creator and can include altering lecture/class delivery methods (e.g., online, recorded, or live), offering online resources (e.g., interactive or downloadable notes, slide decks, extended readings), and changing assignment deadlines, submission styles, duration, and more. While most LMSs offer ease in content presentation and availability, content optimization can prove a more difficult task, since maximizing the user experience will change from student to student.

When using an LMS, leveraging their unique structure when creating content is encouraged, trusting that the instructor can learn how to translate their curriculum into compelling digital modules. Some challenges that arise include writing questions that focus on critical thinking and information recall while not dramatically increasing grading time, designing assessments for open-book and open internet use while not making cheating easier, and verifying that the content offers equal opportunities to all students [6]. Equal opportunity is particularly hard to balance among students since each student will have different accessibility needs, computer quality, and learning patterns. Using the data made available by LMS platforms, instructors can observe student performance patterns and contact individual students to identify possible learning barriers.

In the Canvas LMS, online assessments can be graded either manually or automatically. The structure that allows automatic grading poses limits to the questions types that can be included in the assessment (e.g., multiple choice, true/false, fill-in-the-blank). Canvas can use binary grading to look at whether the input is exactly correct or incorrect with limited interpretive ability such as letter case, decimal numbers, and answer spacing/punctuation. Manual grading opens up other questions types such as open-ended questions including essay entries [7]. While manually-graded assessments take more time and attention to grade, they allow for more variance in student responses and awarding of partial-credit [8]. Interpretation of open-ended student work is a critical area that LMS automatic grading is not geared towards, but the services LMSs do offer can help offset the personal attention needed to assess overall performance.

When making the switch to digital assessments, cheating is a major concern. Some students post questions online and seek external help on assessments. Misrepresentation of student knowledge when online resources are available to students during assessments is not a new phenomena. A study conducted in 2005 compared scores of students who participated in a proctored exam versus an online non-proctored exam where online resources were potentially available but discouraged. Results showed that the online sections scored an average of 10 points more than the proctored sections displaying the advantage given by their exam format [9], [10]. Anti-cheating technology such as lock-down browsers, artificial intelligence proctoring systems, and online proctoring systems can help minimize this type of behavior by using the students' online activity, webcams, and/or microphones to detect suspicious activity [11], [12]. Unfortunately, these anti-cheating measures are imperfect, may send an unintentional message to students, and affect the culture in the classroom.

In an increasingly digital world, we sought to build a different type of online assessment that uses the features of the Canvas LMS to build a unique style of exam. These assessments were timed and allowed students up to 10 attempts to complete them over the course of a two day window. In the C++, Matlab, Engineering Design, and Ethics/IP assessments, there would be 4-5 question banks that each provide one question chosen at random from the bank for each assessment attempt. In the AutoCAD and Solidworks assessments, there is one question bank that randomly determines what figure the student will draw during that assessment. Consistent with online exam best practices [13], these questions were open-ended, difficult-to-lookup questions that focus on completing an original task (e.g., build a novel program, draw a unique part). Although not the focus of this paper, our method seeks to undercut cheating in digital assessments through lowering the anxiety that inspires some students to seek outside assistance, randomizing timed-questions so that cheating methods that take longer will not be useful, and writing questions that limit techniques to those covered in class to make answering these questions correctly more difficult for outside experts with no experience in this particular class.

#### **Motivation and Methods:**

While translating class material to a digital space using an LMS, we sought to create assessments that simulate the timed pressurized environment of an in-class exam that include flexibility for students in different time zones, family needs, and computer issues. Since students are no longer in a shared space, variables such as internet speed/reliability and computer quality can play a large role in the success of individuals. To account for this risk, the assessments in this study were designed to give students greater flexibility while mimicking the design of in-class quizzes and exams. Students were given up to 10 attempts for each assessment with each attempt limited to the approximate length of one class-period to simulate the time-frame of a traditional quiz or exam. The assessments reference a rotating question bank so students would encounter a unique combination of questions when they started each new attempt. Without rotating questions,

students could use two attempts to effectively have the whole two day window to complete the assessment.

Assessments covered C++, Matlab, Engineering Design, Ethics & Intellectual Property (IP), AutoCAD, and Solidworks. Students were given between 100-160 minutes for each attempt, the allotted time being unique to each assessment type. Since there are a limited number of questions in each question bank, a student could take the exam repeatedly until they have completed every possible question and guarantee a high grade. Using this technique, the student would learn more from completing every possible version of the exam, so earning a high grade through high effort is an acceptable method. In addition, taking several attempts in a row to avoid an undesired question is an acceptable test-taking approach as long as they are able to complete the chosen attempt in one iteration before running out of attempts. This framework is designed to give students flexibility to figure out an approach that works best for them. Mirroring the real world, picking your projects carefully can determine your success in a given company or career. Here, we offer students the opportunity to learn how to determine their own success through their choices and effort.

Throughout the different assessments, we used Canvas's 'essay'-type questions frequently. Essay questions give the student a block of space to write free-response answers. In addition, students can make tables, upload files (including images), write sentences, and list bullet points. When grading, the uploaded files in an essay question can be previewed in the browser, unlike files uploaded with the 'file upload'-type questions.

Over the course of three semesters, two cohorts of students used the new assessments and both were asked to take a one-time survey to gather feedback on the assessment style employed through the LMS. The first cohort was the same in the fall and spring semesters. The survey asked the following two questions, one 5-point likert scale (Q1) and one open-ended (Q2):

Q1) I liked this exam policy more than taking an exam once during class time. (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree)

Q2) This year, my exams allowed you to take them up to 10 times in a 2 day window (without grading until the end) with rotating questions using a question bank. What did you think of this policy? How did it affect your approach to exams?

The results for the Q1 were averaged across a 5-point scale. Q2 was analyzed using specific keywords from the responses to indicate positivity, lower stress/anxiety, and flexibility provided by the new assessment structure. Each response in Q2 could count towards 0, 1, 2, or 3 categories if the response included at least one of the identified phrases for a given category. The

total number of matches per category was divided by the total number or responses from students to obtain a percentage of students that mentioned keywords in each of the three categories.

Canvas records how many attempts students used per assessment. By analyzing this data from the LMS, we determined what percentage of students used more than one attempt on each assessment. This data was sorted by the order the assessments were given in the course and color coded by topic.

#### **Results:**

Over the course of 10 unique assessments and 3 semesters, 54.7% of students used more than one attempt on each quiz or exam when averaged across all assessments in the study. Depending on the type of assessment, the percentage of students who used multiple attempts shifted as shown on **Figure 1**. When the student group from the first academic year (Fall 2020, Spring 2021) were newly introduced to this assessment style, they were hesitant to use the additional attempts when not absolutely necessary (e.g., ran out of time on the first attempt, computer problems). This phenomenon is identified by usage rates below 50% during the first 3 assessments. After looking at the low-usage data, the instructor talked to the students about the various ways to use multiple attempts such as using it as a study tool to preview the types of questions that would be asked and using an additional attempt to circumvent a question they were not confident with. After helping to reassure students that using extra attempts in this way was expected and acceptable (and not taking advantage of the system), more students used multiple attempts in their following assessments.

There are a few different ways students can use the multiple attempts to gain an ethical advantage in taking the assessment. If the student solves a question but runs out of time on the exam, they may cycle through several new attempts to find a version of the assessment with the same question and continue to build upon their first attempt. If the student opens the exam and has trouble understanding one or more questions, they may skip through until they find an attempt with questions they feel more confident in solving. Lastly, they might open the assessment early in the 2-day exam window and read through it without attempting to solve any questions to understand what type of questions will be on the assessment and use it as a study tool. Using multiple attempts in creative ways is encouraged to teach students to think about how to effectively use their resources to maximize their goals (usually their grade). Creative ways to use the multiple attempts are discussed in class to encourage their use and emphasize that using the attempts in a variety of ways is acceptable.

Over the course of the first first, students used multiple attempts more often as the year went on as shown in **Figure 1**. There are two notable exceptions to this increased-usage trend in the first year, the Solidworks Quiz 1 and the Ethics/Intellectual Property (IP) Exam. The Solidworks

exam allowed three class periods worth of time per assessment attempt due to the complexity of the objects the students were asked to model and to avoid time pressure in that particular exam. The increase in examination length allowed a higher percentage of students to finish on the first attempt. The Ethics/IP Exam happened during a week where most students had a high number of assignments and assessments for other classes and many felt they understood the concepts well enough that they did not take the exam a second time. In addition, the 2020 Fall AutoCAD Quiz 1 was a take home assessment, so it was not included in this analysis during the 2020 academic year.



**Fig. 1.** Chronological list of the percent of students who used more than one attempt on a specified assessment across three semesters: Fall 2020 (n=123), Spring 2021 (n=117), and Fall 2021 (n=97). The chronological order shows that student usage plateaus around 60% by the end of each term. The assessments are color coded by type to allow comparison of different assessments in the same topic: C++ (light blue), Matlab (dark blue), Engineering Design (light green), Ethics/Intellectual Property (dark green), AutoCAD (light purple).

During the third semester which observed the second cohort of students (Fall 2021), the instructor included an introduction to the ways that multiple attempts can be used alongside the initial introduction to the new assessment style. The combination of communicating both the mechanics and best practices of the system resulted in increased usage of multiple attempts in the first two assessments when compared to the first academic year for the same assessments. After students adjusted to the new system, the usage returned to the low 60% level seen in the previous year.

Coding assessments generally saw the highest usage of multiple attempts. In C++ and Matlab assessments, students can compile and run their code to demonstrate whether they wrote a solution that fulfilled the assessment requirements. Incorrect solutions may take too long to debug successfully, forcing the student to take another attempt because they could not compile their code or tested for errors in functionality. With AutoCAD and Solidworks assessments, students must analyze their own drawings or models to verify their accuracy, but this visual accuracy check is much faster so errors might be fixed during the testing time limit. In the Engineering Design and Ethics/IP assessments, the questions are very open-ended, so students must rely on their own judgment to know when their answer is satisfactory since there is no program to check their work. We speculate that immediate feedback indicating a wrong answer would prompt more students to make another attempt more than quick or ambiguous verification of accuracy.

The students were asked to participate in a voluntary survey at the end of the semester to gauge their opinions on the multi-attempt approach to assessments. Multi-attempt assessments were a success from the students' perspective. On the Likert scale, the students rated the assessments as a 4.79 out of 5 with over 80% of students selecting "Strongly agree" to Q1, affirming that they liked the new assessments more than in-class exams (n=206).

We identified three trends seen across students' responses: positivity, lower stress or anxiety, and flexibility that the new exams provided. Qualitative analysis was used to identify keywords from surveys since overall messages may be relayed through lengthy strings of text rather than a handful of single words. The keywords used in our analysis can be found in **Table 1**.

**Table 1.** Open-response student feedback questions were searched for keywords to identify student opinions of the new assessment method in three categories: Positivity, Lower Stress/Anxiety, and Flexibility. The Positivity keywords are words and phrases that showed that the student 'liked' the new exam system or discussed it with positive adjectives (amazing, enjoyed, nice, great, etc.). The Lower Stress/Anxiety keywords are phrases that specifically mention stress or anxiety, or phrases that discuss comfort and preparedness for the exams. Flexibility keywords express the students' ability to choose when they take their exam looking at conditions such as time zone differences and personal scheduling conflicts.

Student Feedback Survey Response Keywords		
Positivity	Lower Stress/Anxiety	Flexibility
like/enjoy/great appreciate/comforting nice/good policy helped/fan of worked well genius idea amazing/excellent love(d)/better beneficial	lower anxiety lower stress lower/take off pressure relaxed without anxiety peace of mind comfortable calmer less nervous	flexible/flexibility time that works best around my own schedule convenient time zone (difference) choose when to take personal/technical issues

Three methods of categorization were used when interpreting the keyword phrases; perfect matches, interpreted matches, and matching the meaning of a phrase. The three student quotes below illustrate a perfect match, an interpreted match, and a response that could be interpreted to match a keyword but is unclear as a stand-alone response. In each response, the highlighting color code matches the category titles in **Table 1**.

- "I also really enjoyed this policy, as it granted some much-needed flexibility in my schedule, especially as the semester wore on. I'd always give each attempt my all, and knowing that I could use another attempt if need be lessened some of the stress, which helped my performance. Although I rarely needed extra attempts, the fact that they were always there was comforting."
- 2) "Even though I only used 1 attempt pretty much every time, I really liked the option of being able to take it multiple times. This made me feel like if I got stuck or if for any reason I was unable to finish for any reason (i.e. computer problems) I had this sort of safety net that took away most of the stress associated with quizzes/tests."
- 3) "I personally really like this policy or way of administrating the quizzes/exams. I feel like I wasn't rushed and I was able to familiarize myself with the material in the quizzes or exams. It gave me a chance to review things that I didn't remember so there weren't any sudden curve balls that got me stressed out or anything."

Student Quote 1 shows precise keyword matches between the student response and **Table 1**. There is no interpretation needed in order to categorize the quote by response-type. Student Quote 2 shows a phrase 'computer problems' that was interpreted to mean "*technical issues*"

which falls into the code category "*Flexibility*." Student Quote 3 shows where there is a possible argument for inclusion in the "*Flexibility*" category, but 'not being rushed' is not a direct link to the key words associated with the category. The keyword 'take my time' was removed since there is still a test time limit and ultimately, one entire iteration of the assessment must be completed in a timeframe similar to that of a traditional proctored, in-person exam. Therefore, Student Quote 3 was not included in our "*Flexibility*" category.

Qualitative analysis showed that 86.2% of students reported positivity keywords in their responses, 51.2% reported lowered stress or anxiety keywords, and 22.6% mentioned favoring the assessment model's flexibility as shown on **Figure 2**. This is a wonderful early result since the new assessments made a positive impact on so many students. Given the open-ended question did not specifically ask about anxiety, stress, or flexibility, students offered this feedback without prompting.



**Fig. 2.** Analysis of keywords in open-ended student survey responses broken into three categories representing what percentage of total students (n=206) identified with positivity, lower stress/anxiety, and flexibility for multi-attempt assessments. Each category represents a percentage of students who mentioned at least one keyword in their response. The keyword identifiers are noted in **Table 1**. Over 80% of students reported consistent, positive feedback and over 50% of students reported lowered stress or anxiety in regards to assessment delivery over the course of the three semester-long periods.

#### Limitations:

One challenge observed with multiple attempt exams occurs when a student continuously takes the exam until they think that they have exactly 100%. These students end up spending many more hours than necessary taking the exam in an attempt to guarantee a higher grade than their classmates. Often, they will have already achieved a 93+% on one of their first few attempts, but continue working. Despite the negative results such as increased time expense and decreased yield per hour spent, these students have an important lesson to learn from these exams: 'Do not let perfect be the enemy of good'. Learning time management is an extremely important skill to gain early in their university experience and multiple attempt exams provide the opportunity to learn this skill in a low risk environment.

Another potential problem with this technique is vulnerability to students who are determined to cheat their way through the course. There is nothing inherent in this technique that prevents students from talking to one another during the two day window or preventing students from posting questions online. During the first year, we found several questions online in expert-answer-queries after the semester had ended (we were not vigilant enough in checking during the year given the challenges of teaching during the pandemic). During the second year, we changed the questions and highlighted that we were actively monitoring their presence online. As a result, no further questions were posted online. In addition, more time was spent verifying student work and checking for cheating. No further signs of unsanctioned activity were discovered during the semester.

#### **Conclusions:**

Multiple-attempt digital assessments with rotating questions lower the stress students feel when taking an assessment. Throughout three semesters where multiple attempts were used, 86% of students had positive feedback for the new style of assessments. After getting used to the system,  $\sim$ 60% of students used more than 1 attempt on each assessment. Given the positive reception, we plan to continue using this system and refining it in the years to come.

Although successful in this use case, instructors interested in implementing multi-attempt assessments should consider the fairness of their question banks. Each question bank needs to have questions that are a similar level of difficulty and cannot have answers that can easily be found online. We found 3-4 questions per bank works best. With more questions, it becomes harder to grade and maintain the balance of question difficulty. With fewer questions, each attempt feels too similar to one another. Consider adjusting the number of attempts to tune the process to your class. We recommend the 'essay'-type questions to give students the most flexibility when answering and making your grading faster due to preview settings in the Canvas

LMS. As with all digital exams, checking if questions get posted to the internet is an important anti cheating measure.

In the future, we will continue to innovate questions of similar difficulty to populate our question banks and replace questions that finished serving their purpose. Cheating is one weakness identified within this type of exam, so methods to reduce cheating could be explored despite seeing no obvious cheating attempts once we made the students aware of our efforts to combat it. Questions that use student information to change the questions asked are currently being explored (such as a length in inches that is numerically set equal to a birth month or name length). Finally, this style of assessment could be tested in various departments and class levels to explore whether students react similarly across the university ecosystem.

#### Appendix

#### Sample Question Banks - AutoCAD Quiz 1 Fall 2021

Three Duplo blocks and one ruler were given to the students prior to the assessment.

<u>Deliverable</u>: Complete a 3-view (front, top, and side) fully dimensioned, labeled orthogonal drawing of the Combination of Legos pieces shown in the picture below. The front, top, and right configuration must match the picture. Submit a screenshot <u>AND</u> AutoCAD drawing file to under the question by the end of the exam (use the the upload image for the screenshot (circled in red below) and upload documents (2 buttons to the right of the red circle) to upload the drawing files). To see the upload buttons, you may need to click the three vertical dots on the right side of the toolbar.



#### Sample Question Banks - 'for loops' from C++ Exam Fall 2021

- Write a for loop that prints to the screen all of the integers that are divisible by 3 but not divisible by 7 starting at 3 and stopping at 400 (inclusive). Print them in statements like "Next integer is: ", on separate lines.
- 2) Write a **for** loop that checks even integers 4-500 (inclusive) and outputs the number if it is evenly divisible by 4 and 11. Print each integer that meets the conditions on its own line in statements like "Next integer is: ".
- **3)** Write a **for** loop that outputs the average of 4 random numbers (From 20-80 inclusive). Then write another for loop that outputs the average of 30 random numbers (from 20-80 inclusive) and tell the user if the average of the first loop is smaller, larger or remains the same compared to the average in the second for loop.

#### References

- [1] A. Chirumamilla and G. Sindre, "E-exams in Norwegian higher education: Vendors and managers views on requirements in a digital ecosystem perspective," *Comput. Educ.*, vol. 172, p. 104263, 2021.
- [2] A. Manteghinejad, "Web-Based Medical Examinations During the COVID-19 Era: Reconsidering Learning as the Main Goal of Examination," *JMIR Med. Educ.*, vol. 7, no. 3, p. e25355, 2021.
- [3] M. Lapp, J. Ringenberg, and T. J. Fleszar, "Engineering Online Gateway System Ensuring And Evaluating Student Learning Through Automated, Milestone Exams," in *2010 ASEE Annual Conference & Exposition*, 2010, p. 15.495.1-15.495.10.
- [4] A. Mehrabian, T. Ali, and A. Rahrooh, "Crafting Online Exams In Engineering And Technology: Latest Challenges, Methodologies, And Trends," in 2008 Annual Conference & Exposition, 2008, p. 13.338.1-13.338.7.
- [5] D. Draskovic, M. Misic, and Z. Stanisavljevic, "Transition from traditional to LMS supported examining: A case study in computer engineering," *Comput. Appl. Eng. Educ.*, vol. 24, no. 5, pp. 775–786, 2016.
- [6] S. C. Maroo, "Positive Statistical Impact of Online Homework Assignments on Exam and Overall Course Grades," presented at the 2019 ASEE Annual Conference & Exposition, 2019.
- J. Riera, M. Ardid, J. A. Gomez-Tejedor, A. Vidaurre, and J. M. Meseguer-Duenas, "Students' Perception of Auto-Scored Online Exams in Blended Assessment: Feedback for Improvement," *Educ. XX1*, vol. 21, no. 2, pp. 79–103, 2018.
- [8] E. Gehringer, "Online Vs. On Paper Exams," in 2010 ASEE Annual Conference & *Exposition*, 2010, p. 15.927.1-19.
- [9] E. F. Gehringer, A. Menon, and G. Wang, "Tools for Detecting Plagiarism in Online Exams," presented at the 2021 ASEE Virtual Annual Conference, 2021.
- [10] O. R. Harmon and J. Lambrinos, "Are Online Exams an Invitation to Cheat?," J. Econ. Educ., vol. 39, no. 2, pp. 116–125, 2008.
- [11] A. Nigam, R. Pasricha, T. Singh, and P. Churi, "A Systematic Review on AI-based Proctoring Systems: Past, Present and Future," *Educ. Inf. Technol.*, vol. 26, no. 5, pp. 6421–6445, 2021.
- [12] K. Hylton, Y. Levy, and L. P. Dringus, "Utilizing webcam-based proctoring to deter misconduct in online exams," *Comput. Educ.*, vol. 92–93, pp. 53–63, 2016.
- [13] B. Christe, "Writing Online Exam Questions That Discourage Dishonesty," in 2003 ASEE Annual Conference, 2003, p. 8.1321.1–6.