The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #43404

Paper or Silicon: Assessing Student Understanding in a Computer-based Testing Environment Using PrairieLearn

Mr. Jamal Ardister, Michigan State University
Dr. Geoffrey Recktenwald, Michigan State University

Geoff Recktenwald is a member of the teaching faculty in the Department of Mechanical Engineering at Michigan State University. Geoff holds a PhD in Theoretical and Applied Mechanics from Cornell University and Bachelor degrees in Mechanical Engineering

Sara Roccabianca, Michigan State University

Sara Roccabianca is an Associate Professor in the Department of Mechanical Engineering at Michigan State University (MSU). She was born and raised in Verona, Italy and received her B.S. and M.S. in Civil Engineering from the University of Trento, Italy. S

Paper or Silicon: Assessing Student Understanding in a Computer-based Testing Environment using PrairieLearn

Abstract

Computer-based testing is a powerful tool for scaling exams in large lecture classes. The decision to adopt computer-based testing is typically framed as a tradeoff in terms of time; time saved by auto-grading is reallocated as time spent developing problem pools, but with significant time savings. This paper seeks to examine the tradeoff in terms of accuracy in measuring student understanding.

While some exams (e.g., multiple choice) are readily portable to a computer-based format, adequately porting other exam types (e.g., drawings like FBDs or worked problems) can be challenging. A key component of this challenge is to ask "What is the exam actually able to measure?" In this paper the authors will provide a quantitative and qualitative analysis of student understanding measurements via computer-based testing in a sophomore level Solid Mechanics course.

At Michigan State University, Solid Mechanics is taught using the SMART methodology. SMART stands for Supported Mastery Assessment through Repeated Testing. In a typical semester, students are given 5 exams that test their understanding of the material. Each exam is graded using the SMART rubric which awards full points for the correct answer, some percentage for non-conceptual errors, and zero points for a solution that has a conceptual error. Every exam is divided into four sections; concept, simple, average, and challenge. Each exam has at least one retake opportunity, for a total of 10 written tests.

In the current study, students representing 10% of the class took half of each exam in Prairie Learn, a computer-based auto-grading platform. During this exam, students were given instant feedback on submitted answers (correct or incorrect) and given an opportunity to identify their mistakes and resubmit their work. Students were provided with scratch paper to set up the problem and work out solutions. After the exam, the paper-based work was compared with the computer submitted answers.

This paper examines what types of mistakes (conceptual and non-conceptual) students were able to correct when feedback was provided. The answer is dependent on the type and difficulty of the problem. The analysis also examines whether students taking the computer-based test performed at the same level as their peers who took the paper-based exams. Additionally, student feedback is provided and discussed.

Introduction

Automated grading has been around since Michael Sokolski invented scantron grading machines in 1972. Over time, computers have evolved from grading multiple choice exams to accepting numerical and written solutions. New systems like PrairieLearn can grade a wide variety of solutions, including engineering sketches like Free Body Diagrams. This paper approaches the idea of using automated grading in conjunction with the SMART pedagogical methodology.

SMART

The supported mastery assessment through repeated testing (SMART) model discourages ineffective studying habits such as problem memorization and copying of homework solutions from various sources such as online sources, solution manuals, and friends [1]. Not only does it discourage bad learning habits, it has also been shown to improve student understanding and problem-solving ability by encouraging students to better understand theory and concepts which can be seen through help room and office hours interactions with students [2,3]. While some course dependent modifications may be required, it has been shown that the method can be adapted to suite other classes as well [4]. Additionally, small case studies have shown that the SMART model is not only effective in the class as a whole. These results have been shown to be consistent for underrepresented minorities and women as well [5].

One motivating factor for the development of the SMART method is the observation that students often use ineffective approaches to learning material. Problem memorization allows the student to imitate conceptual understanding and consequently receive partial credit on exams without understanding the material and/or retaining misunderstandings. This behavior may cause downstream difficulties in later classes as new material is built on a non-existent foundation. The SMART model is designed to minimize the efficacy of the bad study habits by requiring students to systematically solve problems and demonstrate mastery. In the SMART grading process shown in Table 1, full credit (100%) is awarded if and only if students obtain the correct answer with

Table 1 – SMART method grading rubric. The following table shows the well-defined partial credit rubric used in the SMART method.

Competency	Level	Score	Description
Meets Minimum Competency	I	100%	Correct answer fully supported by a complete, rational and easy to follow solution process, including required diagrams and figures
	II	80%	Incorrect answer due to one or two minor errors but supported by a correct solution process as described in Level I
Does Not Meet Minimum Competency	III	0%	Incorrect answer due to conceptual error(s)

reasonable support. Partial credit (80%) is only given for non-conceptual errors. Non-conceptual errors include calculation errors or algebraic errors as well as other simple mistakes. The opposite, a conceptual error would include mistakes such as invalid free body diagrams, missing components of a stress state, sign errors in moment calculations, or equations that are inconsistent with a free body diagram or coordinate system. The focus on conceptual understanding awards points only when students clearly achieve the student learning outcomes. The clear consistent use of partial credit makes it much harder to "game" or "cheat" than traditional scoring methods. Because students are motivated by 'the grade', they adopt study habits that lead to long term learning and achieve competency with the material.

The SMART model also encourages students to develop troubleshooting skills. When exams are graded each problem is awarded either 100% for correct or 0% for incorrect regardless of the error. The burden of obtaining partial credit lies on the student. The student must present a written appeal for partial credit. To write the appeal, the student must review their work and find their mistake. The appeal must contain a description of the mistake and a rework of the problem. The appeal must explain where the error was made and show that it was a non-conceptual error. Appeals are reviewed by the instructor or teaching assistant, and the grade may be changed to 80% if the error is truly non-conceptual. Students have a second chance to take an exam on similar content the following week to demonstrate mastery of the concepts.

The RT in SMART stands for repeated testing. This two-step exam process happens every three weeks, resulting in 10 exams over the course of the semester. As such, there is a significant but not unmanageable amount of exam writing, exam grading, and appeal grading. Historically, this process has been beneficial, but frequency of exams could be a barrier to adoption of the SMART method for both students and instructors.

A potential solution to this challenge is automating the grading process of the first pass grading (0%, 100%). If automatic grading could also identify conceptual and non-conceptual errors, it could be used to process the appeals as well. Resulting in a significant reduction in grading and a faster turnaround time for getting grades and appeals back to students.

Computer-Based Testing

Computer based testing has been shown to require a large "upfront" time commitment but then requiring much less time to maintain [6]. In the same study [6], the authors also showed that the benefits increased proportional to number of students in the class (exams needing to be graded and time required to grade them). Without teaching approaches that can scale with size, the effort it takes to grade large classes (~200 students) can be onerous. Looking at the results of the computer-based exams, the effort remains constant and does not grow with class size since the computer auto-generates and auto-grades the exams [6].

In this work the authors chose to utilize a computer-based testing system known as PrairieLearn, an automated grading system unique in its flexibility and ability to accept answers in a broad array of formats [7]. Problems are written in HTML and supported by a Python script to handle variable or even image randomization. For example, an instructor may wish to write a problem that evaluates a student's ability to calculate a moment. The problem, written in terms of its parameters, would cycle through many variants while maintaining the core conceptual question. The problem configuration in computer-based testing also allows for a wide range of options. Problems can be configured to set the number of attempts, value of each attempt, error tolerance, etc. This level of customization is ideal for engineering faculty who typically require graphical, vector, or unit-based answers. Computer-based testing systems can also be designed to accept many types of answer inputs such as button selection, text and/or numerical input (with or without units), units only, graphical drawings involving sketching functions (like a shear and moment diagram) or placing symbols (like a Free Body Diagram or FBD). We chose to use PrairieLearn due to its versatility. In this study, the authors do not employ the randomized parameter values so the test is identical to the paper version taken by the rest of the class, thus isolating the effects of implementing SMART while using a computer for instant feedback. Future studies will utilize the randomization feature.

Purpose

The purpose of this paper is to examine the effectiveness of using a computer-based testing environment or system, such as PrairieLearn, to grade exams consistent with SMART pedagogy and thus show that the SMART model can be employed by means of computer-based testing.

Methods

During the fall semester of 2023, students in ME222 were given the opportunity to take exams in a paper format or on a computer (with full paper exam provided). The students took the computer-based exam in a proctored computer lab at the same time as the rest of the class took the regular exam. Both exams were identical on paper, but there were small differences in how the computer-based system requested the answers (See Figure 1 for an example).

Students were given the option to take either exam, so the subset is self-selected, however, it is assumed that this selection method does not impact the results of the study. The students were briefed on the process and how it would not impact their grades. Taking a subset from the entire class allowed us to directly compare students using computer-based testing and those using paper-based testing. This study only covered simple and concept problems. These problems typically have up to three computational steps and are less involved than a textbook problem. The class had an enrollment of 150 students and throughout the semester about 10% of the class, or 15 students, took the optional computer-based exams.

Students are often stressed due to time constraints during an exam. To ensure that this was not an issue, students taking the computer-based exams were allotted an additional 30 mins during the exam to account for the time needed to enter answers and track down mistakes. This had the added benefit of alleviating stress due to computer issues. Importantly, all exams in this class are designed to take less than the allotted time so students have ample space to review their work for errors.

Students taking the computer-based version were told to write their work on the paper as they normally would and then enter their answers into the computer. The computer would auto-grade the answer and provide instantaneous feedback in the form of an "x" or "✓" for each piece of the problem. For the simple problems, students were allowed 4 or 5 attempts to answer the question. The first attempt was worth 100% and each subsequent attempt was worth 80% as seen in Table 2.

Table 2 – Computer based grading rubric – The following table shows how attempts were auto-graded.

Competency	Level	Score	Description
Meets Minimum Competency	I	100%	Correct answer on Attempt #1
	II	80%	Correct answer on Attempts #2 - #5
Does Not Meet Minimum Competency	III	0%	No correct answer

After the exam, the authors reviewed the written work and the computer-based answers (all attempts) to determine what error was made and whether it was conceptual, or non-conceptual. These errors were then categorized according to the ID in Table 4 to evaluate the compatibility of the computer-based rubric and the SMART method rubric. That is, does the computer-based system adequately evaluate answers and auto-grade consistent with the SMART rubric we have been using on paper exams.

Multiple Attempts and Conceptual Understanding

A key aspect of the SMART method is to allow students to find and correct their errors. Thus, computer-based exams should have an option for multiple attempts. However, the number of attempts should be low enough that guessing is discouraged. Recent work by Zilles, et al [8] has shown that students studying for exams benefit from 1-3 attempts, but not much more. After 3 attempts students typically give up [8]. In SMART methodology terms, if students cannot obtain the correct answer in a few tries on a simple or concept problem, they likely have a conceptual error rather than an algebraic error. This self-imposed attempt limit due conceptual misunderstanding can be tested by limiting attempts during an exam on the computer-based exam while comparing the students answers with a handwritten version that they simultaneously work on scratch paper. A student with proper conceptual understanding will likely get the problem correct in 1-3 attempt as in [8] and the scratch work will also show proper understanding. While conversely, a student with a conceptual misunderstanding will likely give up after 1-3 attempts as in [8] with the misunderstanding also showing up in the scratch work.

Selecting Problem Difficulty

In the SMART method exam problems are divided into four basic categories:

1. *Concept Problems* test conceptual understanding or supporting knowledge. A concept problem may be a free body diagram, a description of boundary conditions, or writing a constraint equation. These problems are straight forward or single step expressions of a concept and thus, are typically unappealable.

2. *Simple Problems* test basic skills needed to solve more complex problems. A simple problem will contain a minimal number of steps or computations. Identifying an algebra error in a simple problem should be straightforward because there are minimal steps.

3. *Average Problems* test the student's ability to contextualize a problem and work through multiple steps. Most problems in standard textbooks are at the average level. Average problems may require a page or two of drawn images and algebra. A two-cut shear and moment diagram is a good example of an average problem as well as problems involving two-force members.

4. *Challenge Problems* test the student's ability to apply their problem solving process to a problem with more complexity than an average problem. The more difficult problems in a textbook would be at the challenge level. Problems in 3D or requiring a full stress state are good examples for this category.

When exams are offered in the computer-based testing environment, each of these difficulty levels presents different challenges.

Concept problems

Concept problems could involve drawings and sketches which are difficult to parameterize and tedious to code without error. Concept problems that require students to write an equation could be moved to multiple choice, but there is a big difference between selecting from pre-formed answers and having to develop the equation oneself. That being said, concept problems are the easiest to understand from an auto-grading point assignment perspective. The answers are either right or wrong. Repeat attempts are typically not given unless a new problem is instanced.

One example of a concept problem tested through computer-based testing is interpreting a 2D stress element (Figures 1a and 1b). We make two observations about these problems. First, the sign of the stress is a key item being measured. If a student is given multiple attempts to solve the problem, they might immediately start guessing signs or even permutations. Second, Figure 1b is not easy to set up for auto-grading because it is too open ended. However, Figure 1a gives structural hints to the student since the subscripts and Greek notation is an aspect of what is being tested. One way to test this skill is to insert distractors, like option (c) in Figure 1a.

Section 1: Concept problems [4 pts]

Problem 2. Write the stress state associated with the 2-D stress element shown below. *Pay careful attention to the coordinate system defined and subscripts.*

(a) σ_x = __________

(b) σ_y = __________

(c) σ_z = __________

(d) τ_{yx} = __________

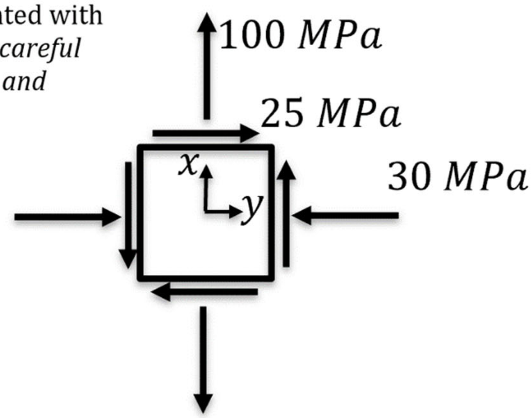

Figure 1a - An example of a concept problem with structural hints.

Section 1: Concept problems [4 pts]

Problem 2. In the box provided, write the stress state associated with the 2-D stress element shown below. *Pay careful attention to the coordinate system defined and subscripts.*

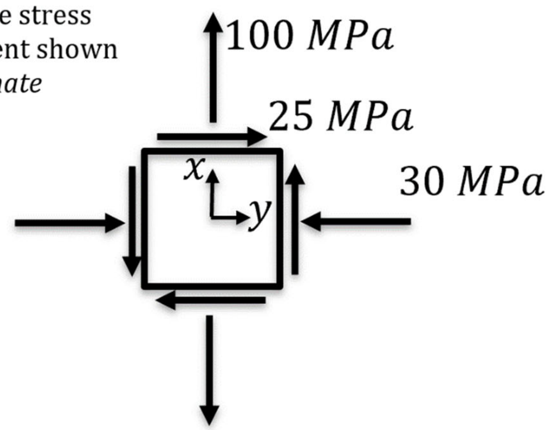

Figure 1b - An example of a concept problem without structural hints.

Simple problems

Simple problems typically have one or two answers and minimal steps. As such, students who have algebraic or other non-conceptual mistakes can often find their errors if they are told that their answer is wrong. In this sense, providing students with instant feedback allows them to 'appeal' their work while the exam is in progress, thus auto-grading at 100% for correct answers on the first attempt, 80% for a correct answer on a later attempt, and 0% after a few attempts. This scheme should be able to accurately capture the SMART ethos. While FBDs are typically required on paper versions of the exam, requiring many FBDs in computer-based testing environments are expensive for both students (time required during the exam) and instructors (time required to create the problem). Thus, the FBDs are tested as concept problems and not tested again for simple problems.

A benefit with the simple problems is that they are short. However, because there are only a few steps, there is a chance that a student could make canceling errors like a double negative that hide conceptual misunderstandings or sloppiness. It is presumed that these mistakes should be rare enough that it does not affect a student enough to change the final grade, but we do measure them in this study. A bigger problem is when students try to guess solutions. A small number of steps means that a student may be able to game the equations by guessing without fully understanding the problem ultimately resulting in a correct answer with a conceptual misunderstanding.

Average and Challenge problems

Average and challenge problems have enough steps that telling a student that the final answer is wrong will likely not result in them finding their mistake. There are just too many places they could have made an error. This can be managed by providing students with the opportunity to validate their intermediate solutions. However, anecdotally, this mid-step verification process can increase anxiety among the students. Although we consider computer-based testing to be useful for average and challenge problems because it allows the first pass grading, review of the handwritten solution is still necessary to identify non-conceptual errors and award points accordingly. This issue will be the scope of a future study.

Classification

The question being asked in this paper is: "how accurately does the computer-based testing/feedback identify student competencies?" As a first step in this process, we classified the possible scenarios and assess how a computer-based testing system and SMART grading would score those scenarios. Table 3 shows a set of student performance scenarios seen or anticipated during this study and classifies them as consistent or inconsistent with the SMART method.

Results

Detailed results are provided for a single problem. Although the results from other problems were not identical, this problem is a reasonable sample of the results seen.

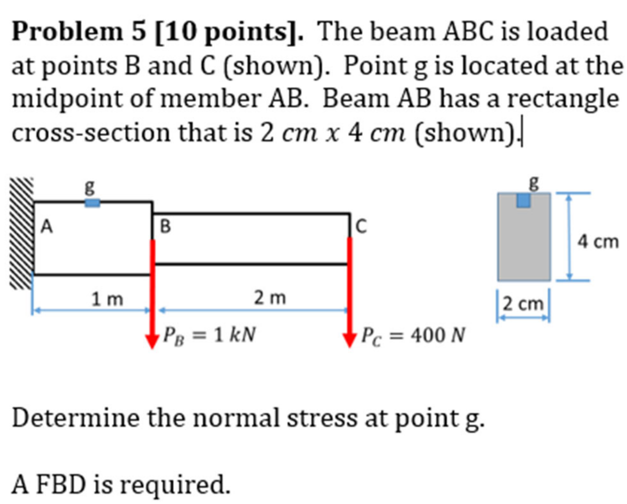

Figure 2 – The bending moment stress problem being analyzed in this section.

Problem 5, shown in Figure 2, is a "simple" problem. It required students to compute the axial stress caused by a bending moment in the beam. The problem is complicated by the presence of two loads at different locations. A typical solution requires a Free Body Diagram with internal reactions based on the right-hand rule coordinate system, a valid set of equilibrium equations, solving for geometric parameters (I_z and y), and using the bending moment stress equations. The correct answer involves three equations (equilibrium for moments, second moment of area, and stress) which can be solved serially to obtain a single value for the stress. Students must also consider appropriate unit conversions.

Table 3 - Student performance scenarios and grading consistency. The ID is provided to connect Table 1 Scenarios with examples of each of these scenarios.

ID	Scenario	Computer-based auto-grading	Paper-based manual-grading	Result	Comments
C1	Student gets the correct answer for the right reasons.	100%	100%	Consistent	
C2	Student makes a non-conceptual mistake, but after feedback corrects the error.	80%	80%	Consistent	
C3	Student makes a conceptual error and never arrives at the correct solution	0%	0%	Consistent	
C4	Student makes a non-conceptual mistake, and after feedback finds the right answer, but for the wrong reasons.	80%	80%	Consistent	While the grading is consistent, the student does not know what they did wrong and may make a similar mistake.
P1	Student makes canceling errors and arrives at the right solution	100%	0%	Inconsistent	Not a likely scenario in difficult problems and may not be caught by a human grader either.

Table 3 (Continued) - Student performance scenarios and grading consistency. The ID is provided to connect Table 1 Scenarios with examples of each of these scenarios.

ID	Scenario	Computer-based auto-grading	Paper-based manual-grading	Result	Comments
P2	Student does not follow a consistent process or makes an error that does not affect this answer.	80%	0%	Inconsistent	Not a likely scenario in difficult problems, but harmful on simple problems if not corrected.
P3	Student makes a conceptual error, but after feedback corrects the error.	80%	0%	Inconsistent	Acceptable inconsistency since it reinforces learning. Paper testing could not do this.
P4	Student makes a conceptual error, but after trial and error guesses the right solution.	80%	0%	Inconsistent	This can occur if the number of attempts is high compared with the difficulty of the problem.
S1	Student has the correct answer for the right reasons but makes an input error that is or is not corrected.	80% or 0%	100%	Inconsistent (fixable)	Requires faculty intervention, but since PrairieLearn records all student responses, can be done.
S2	Student has the correct answer for the right reasons but uses a different coordinate system or sign convention and is marked wrong.	80% or 0%	100%	Inconsistent (fixable)	Requires that the student's scratchwork is collected and can be associated with the exam.

Table 4, below, shows the mistakes that students made, the outcome (did they recover and eventually get the answer or not) and how the computer-based system and SMART graded the effort. The “Key Mistake” column briefly describes the main error that the student made. The ID category connects the scenarios from Table 1 with these results. ID’s starting with C are consistent between auto-grading and manual-grading.

Table 4 - Grading comparison for the computer-based system and SMART for problem 5 of the exam for each of the 16 students in the computer-based exam.

Student #	Key Mistake	Outcome	Computer-based auto-grading	Paper-based manual-grading	ID Category
1	None / Valid Answer.		100%	100%	C1
2	Minor error, introduced a factor of 2.	Corrected on 2nd attempt for 80%	80%	80%	C2
3	Decimal and algebraic sign errors.	Corrected both but used a few tries to guess the sign.	80%	80%	C2
4	Unit Conversion and unit syntax (mPa vs Mpa).	Corrected both.	80%	80%	C2
5	Moment taken at the wrong location and invalid FBD.	Not recovered.	0%	0%	C3
6	Started, but didn't complete.		0%	0%	C3
7	Moment taken at the wrong location and invalid FBD.	Not recovered.	0%	0%	C3
8	Invalid Moment equation.	Not recovered.	0%	0%	C3

Table 4 (Continued) - Grading comparison for the computer-based system and SMART for problem 5 of the exam for each of the 16 students in the computer-based exam.

Student #	Key Mistake	Outcome	Computer-based auto-grading	Paper-based manual-grading	ID Category
9	Started, but didn't complete.		0%	0%	C3
10	Moment arm issues.	Not recovered.	0%	0%	C3
11	Invalid approach.	Not recovered.	0%	0%	C3
12	Sign error resulting from missing =0 in equilibrium equation.	Guessed by changing a sign but didn't know why and changed a sign at a new place resulting in 2 errors.	80%	80%	C4
13	Student drew the FBD incorrectly and made a sign error in the moment equation.	These two mistakes resulted in a correct answer.	100%	0%	P1
14	FBD missing internal forces and coordinate system.	Since the axial force was not needed to solve the problem, PrairieLearn awarded full points.	100%	0%	P2
15	Sign error in the moment equation.	Guess the right sign.	80%	0%	P4
16	Very haphazard work, missing equilibrium equations that result in sign errors.	Guessed by changing sign.	80%	0%	P4

In this particular problem 12 of the 16 students or 75% had compatible scores (C1-C4). Of the four incompatible scores, a harder problem or requirement to provide an intermediate step answer in computer-based system may have prevented the incompatibility.

A second question is how well did the students taking the computer-based exam perform relative to the rest of the class. Table 5, below, compares the results of the computer-based students with the rest of the class. The middle two columns summarize the results from Table 4, the last column shows the results from the classroom-based exam.

Table 5 - Comparison between grades for students who took the computer-based exam compared to those who took the in-person exam. Computer-based grades for both methods (SMART and PrairieLearn) are provided. The number preceding the percentile is the number of students in that category. The percentile is associated with the group opposed to the whole.

Problem Score	Computer-based exam with manual-grading of written work	Computer-based exam with auto-grading of entries	In class, paper exam with manual-grading of written work
100%	1 (6%)	3 (19%)	38 (27%)
80%	4 (25%)	6 (38%)	14 (10%)
0%	11 (69%)	7 (44%)	89 (63%)

Students taking the computer-based exam may have been more willing to test an answer out without double checking their work, resulting in fewer 100% grades. However, they seemed more able to find mistakes and correct them than the paper students. This held regardless of grading method. This result is seen in other work examining how students perform when given multiple attempts on exams [8].

Student Feedback

Students taking the computer-based exam were given the option to take an informal survey to assess their preferences and how they approached the exam. On a 1-5 Likert scale with 5 being the highest, 5 out of 8 respondents indicated they were very satisfied with the policies (Fig. 3) and 8 out of 8 respondents preferred the computer-based exams over paper-based exams.

Students were also asked questions about how they engaged with the test after getting feedback (Fig. 4). All students indicated they “Always looked for a mistake and tried again”. Students also indicated that they would sometimes “try a guess”, but would often “try the opposite sign”. Students were allowed to move forward and backwards through the exam and most indicated that they used this policy to their advantage.

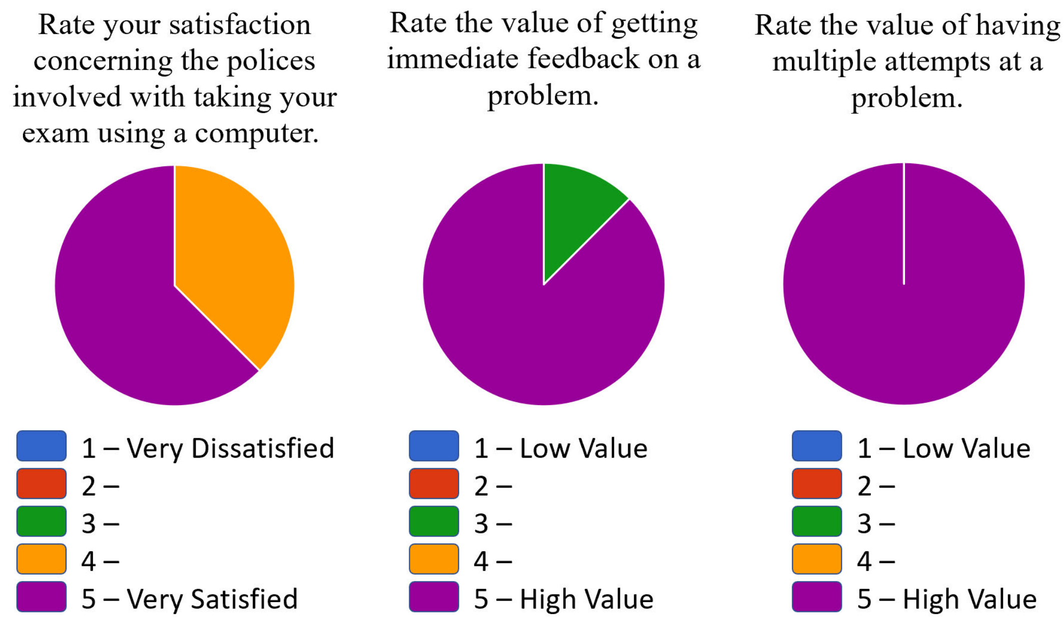

Figure 3 – Survey responses on preferences – Eight students responded to our voluntary survey. Satisfaction with the computer-based testing policies was high and students indicated they highly value multiple attempts and instant feedback.

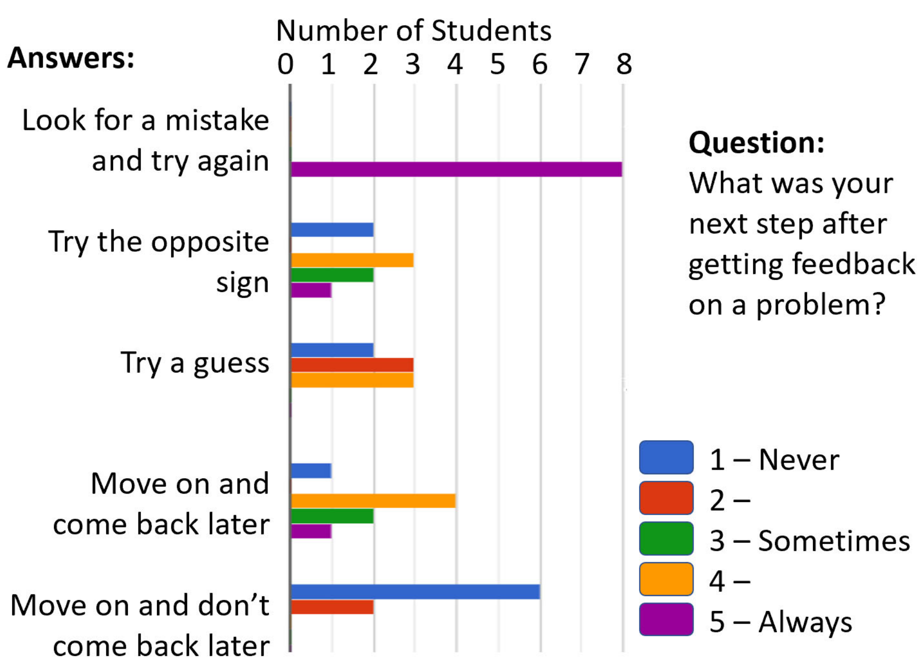

Figure 4 – Survey responses on methods – Eight students responded to our voluntary survey. These answers give insight into how students respond after getting instant feedback that their answer was incorrect.

Informal discussions with students provided some additional insights. Upon completion of proctoring an exam, informal discussions with students supported the idea that students valued the auto-graded exams over manual-graded exams because of the instant feedback. Students express their willingness to take exams using the instant feedback of the computer-based system in conjunction with their normal SMART exam because they feel it allows them to focus on individual questions rather than the exam in its entirety. One student in particular has stated that while the instant feedback can be jarring at first, the reward for knowing you got the answer right far outweighs the negative of seeing that you got the answer wrong. The student explained that knowing the answer was correct allowed for a complete focus shift because that problem was no longer a concern. In essence, the student conveyed that it was less stressful which made better focus and a higher grade possible.

Discussion

A recent study has shown that students can achieve a higher score on an exam or assessment given the opportunity to try it again [8] while additionally making students and faculty happier [9]. We have also seen the SMART method yields higher scores on average without curving grades or awarding points to students using ineffective study strategies. Combining the two methods, computer-based testing and SMART has shown us that students benefit from this combination if done carefully and thoughtfully.

Numerically, a computer-based system is an effective replacement for the appeals process and exam grading for simple and concept problems given in the SMART modality provided the exam authors think about the problems carefully. The problem structure and number of attempts should be carefully chosen so that guessing will not lead to correct answers (or hints).

Future Work

While this data is promising, increasing the sample size and number of problems analyzed is an important next step. Additionally, the authors are working to adequately scaffold average and challenge problems without giving hints on the process or cause undue stress if early answers are incorrect.

Acknowledgements

This work was supported by NSF Grant #2013286.

References

[1] Ron Averill, Geoffrey Recktenwald, Sara Roccabianca, and Ricardo Mejia-Alvarez. "The Need for Holistic Implementation of SMART Assessment". *2020 ASEE North Central Section conference, Morgantown, West Virginia, 2020, March*. ASEE Conferences, 2020.

[2] Ronald C. Averill, Sara Roccabianca and Geoffrey Recktenwald. “A Multi-Instructor Study of Assessment Techniques in Engineering Mechanics Courses.” *Conference Proceedings of ASEE Annual Conference & Exposition*, Tampa, Florida, June 16-19, 2019.

[3] Hjelmstad, K. D., & Baisley, A., “A Novel Approach to Mastery-based Assessment in Sophomore-level Mechanics Courses” *2020 ASEE Virtual Annual Conference,* Virtual, Online, 2020.

[4] Recktenwald, G., & Grimm, M. J., & Averill, R., & Roccabianca, S., “Effects of a New Assessment Model on Female and Underrepresented Minority Students” *2020 ASEE Virtual Annual Conference,* Virtual, Online, 2020.

[5] Recktenwald, G., & Bush, T., & Averill, R. (2020, June), “A New Assessment Model, Modified for Use in Dynamics” *2020 ASEE Virtual Annual Conference,* Virtual, Online, 2020.

[7] West, M., & Herman, G. L., & Zilles, C. “PrairieLearn: Mastery-based Online Problem Solving with Adaptive Scoring and Recommendations Driven by Machine Learning Paper” *2015 ASEE Annual Conference & Exposition*, Seattle, Washington, 2015.

[8] S. Mahmood, M. Zhao, O. Khan and G. L. Herman, "Caches as an Example of Machine-Gradable Exam Questions for Complex Engineering Systems", *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1-9, 2020.

[9] C. Zilles, M. West, D. Mussulman, C. Sacris., “Student and Instructor Experiences with a Computer-Based Testing Facility”, *EDULEARN18 Proceedings*, pp. 4441-4450, 2018.