

The COVID-19 Pandemic and the Challenges of E-Assessment of Calculus Courses in Higher Education: A Case Study in Saudi Arabia

Fatima M. Azmi* and Heba Bakr Khoshaim

Department of General Sciences, Deanship of Educational Services,
Prince Sultan University, Saudi Arabia
<https://orcid.org/0000-0001-9275-0965>
<https://orcid.org/0000-0002-5898-0991>

Abstract. The COVID-19 pandemic has affected many aspects of our lives, including education. Due to this unexpected catastrophe, education has shifted to virtual-learning and auto-grading models in most parts of the world. This study explores the validity and appropriateness of auto-grading-assessment for online exams by comparing students' online exam scores where they are first auto-graded and then manually graded. Furthermore, it investigates whether the mean differences in their scores are statistically significant. The study included two calculus courses taught by the authors, during the spring semester 2019-2020 at a private university in Saudi Arabia. The online exam was performed on the WebAssign platform, which has built-in calculus questions. The sample consisted of fifty-five students who were registered on those calculus courses. The quantitative data was analysed using the SPSS statistical tool. A paired t-test at an alpha level of 0.05 was performed on differences in mean exam scores between auto-graded and manually-graded scores. The statistical analysis results revealed a statistically significant difference in students' mean scores. Our findings illustrate the importance of human intelligence, its role in assessing students' achievements and understanding of mathematical concepts, and the extent to which instructors can currently rely on auto-grading. A careful manual investigation of auto-graded exams revealed different types of mistakes committed by students. Those mistakes were characterized into two categories: non-mathematical mistakes (related to Platform Design) and minor mathematical mistakes, which might deserve partial credit. The study indicated a need to reform the auto-grading system and provided some suggestions to overcome its setbacks.

Keywords: COVID-19 pandemic; E-assessment; validity of auto-grading; higher education; mathematics

*Corresponding author: Fatima M. Azmi; Email: fazmi@psu.edu.sa

1. Introduction

Pandemics threaten people's existence and health status. The associated damage also affects our economic, social, and educational systems. The most recent pandemic is coronavirus—known as COVID-19. An estimated 300 million students are at home from school due to this unforeseen pandemic (McCarthy, 2020). Despite conflicting arguments about the effectiveness of keeping children at home (Viner et al., 2020), decisions to keep all schools closed were consistent worldwide. Both primary/secondary and university students' educational journey has been unexpectedly and severely disturbed in order to contain the virus. In the Kingdom of Saudi Arabia (KSA), to keep students on track, several electronic platforms have been adopted to switch classes to virtual teaching. Online education has traditionally been regarded as an alternative system of teaching, but due to the COVID19 pandemic, educators and students of all levels primary/secondary and university quickly adopted virtual teaching and learning (Hodges, 2020). Thus, the year 2020 has seen increased adoption of electronic resources which can be integrated into virtual learning, for example Zoom and Google Meet (Fulton, 2020).

Technology undoubtedly brings innovative tools and opportunities to the education field (Broughton et al., 2013; Parshall & Guille, 2015). It can support educators to teach and help students to learn (Raja & Nagasubramani, 2018). Particularly in the situation of COVID-19, experiential learning has led to innovations such as virtual labs and field trips (Pennisi, 2020). However, any opportunity is expected to present some challenges (Azevedo, 2015), especially when assessing mathematical knowledge. For example, students must possess other knowledge and master new skills (when using a computer to submit answers) and a lack thereof will affect their success (Alqahtani & Rajkhan, 2020; Parshall & Guille, 2015). Complexity ensues in terms of additional requirements that will add an extra load onto students: familiarity with the online assessment platform, how to write the mathematical formulae, and how to enter the answers in a proper format.

E-assessment or online assessment (auto-grading) refers to the use of information technology to assess students' performance in online exams or homework. One of the challenges of using E-assessment is the restriction in the format of the items: MCQs (multiple choice questions), true/false, matching, or short answer submission type items. Although such a format is convenient for large-stake exams as it provides easy and consistent grading (Stankous, 2018), it might indicate that only the final answer is what matters. In mathematics education, we do not only consider the final answer; we evaluate the logical thinking and decisions made by the student at each step. Thus, the final answer carries little weight when it comes to mathematical problem assessments. In that, partial scoring is important.

Moreover, MCQ or true/false formats might not be appropriate to assess students' analysis and process skills, which are crucial for mathematical competencies. However, there is an ongoing debate concerning the best item-format that can be used for mathematical assessment (Stankous, 2018). Hence,

there is a need to explore the validity of online auto-grading compared to manual-grading. Previous research has shown inconsistencies between manual-grading and auto-grading (Bejar et al., 2017; Stankous, 2018). Stankous argued that true assessment of mathematical knowledge can only be examined through constructive response items.

1.1 Significance of the Study

The use of E-assessment is inevitably growing every day. Due to the COVID-19 pandemic, education shifted to E-learning and E-assessment modes overnight for all schools and universities in Saudi Arabia. Educators did not get a chance to think, evaluate, or even design their E-assessment approaches. Instead, they were limited to choosing from several available platforms, adapt to online teaching, and immediately apply E-assessment procedures. It is now essential to address this approach, evaluate it, and suggest solutions for its limitations. Especially as the COVID-19 pandemic cases began to rise again worldwide, indicating more reliance on virtual teaching and online exams in the future.

This study involves two calculus courses (Calculus I and Business Calculus), which are taught at a private university in Riyadh, KSA. The instructors of the courses (who are the authors of this paper) utilized the WebAssign platform by Cengage for the E-assessment processes for both courses during the spring semester of the academic year 2019-2020. The sample consisted of a total of fifty-five students who were registered on those two courses. This study addresses the following three questions:

Q1: Was there a significant difference in students' online midterm and final exam scores between auto-graded and then manually graded scores for the Calculus 1 course?

Q2: Was there a significant difference in students' online midterm and final exam scores between auto-graded and then manually graded scores for the Business Calculus (B.C.) course?

Q3: Were there some common mistakes committed by students during the online exams due to their unfamiliarity with the WebAssign platform, and did this result in zero credit when auto-graded?

The article findings illustrate the importance of human intelligence, its role in assessing students' achievements and understanding of mathematical concepts. The results provide some suggestions and guidelines for E-assessment methods to assess undergraduate mathematical skills and the need to make auto-grading more intelligent.

2. Literature Review

2.1. E-Learning & E-Assessment in Mathematics

The use of technology is increasingly common in our lives, and the education system is no exception. E-learning in mathematics classrooms refers to the use of technology in the learning process, which can be through multimedia (e.g., videos) to present and articulate a task, software applications to facilitate understanding of mathematical concepts, or online platforms to practice and visualize the problems. Such software has been shown to support students'

conceptual understanding and the depth of their investigation (Heid, 1988), shift their attention from computation and memorizing formulas to help them focus instead on the mathematics itself (Camacho Machín et al., 2010), and improve their visualization skills (Baki et al., 2011). Therefore, E-learning has become a prevalent method of effective learning (Kerzic et al., 2018). In fact, E-learning has shifted from being luxurious to being essential during the COVID-19 pandemic (Alqahtani & Rajkhan, 2020). Most educational institutions transformed to e-learning mode to control the spread of the virus. However, although educators see it as a promising future for teaching and learning (Alanazi & Alshaalan, 2020) and students consider it a time saving method (Khalil et al., 2020), E-learning is not always the preferred approach by students, especially in applied sciences (Abbasi et al., 2020).

Assessment is defined as "the process of collecting evidence regarding students' learning, interpreting the evidence and defining an action" (Shahbari & Abu-Alhija, 2018, p. 1316). The process itself is valuable and has critical consequences for students' futures. Its validity and appropriateness have been investigated in the literature. Until now, assigning numerical or alphabetical grades has been the most common approach in education to represent students' achievements or knowledge and is the ultimate goal of most educational systems (Rešić et al., 2017). However, the assigned grades are not always satisfiable or agreed upon by students or their families (Rešić & Halilčević, 2014). E-assessment, on the other hand, refers to any assessment done electronically, in fact it is any type of computer-based assessment using a given platform. In recent years, the approach has grown in popularity and is expected to be adopted by higher education institutions in the future.

2.2. Benefits of E-assessment

E-assessment is a crucial aspect of online education. Some of its benefits include its fast scoring and cheap administration (Broughton et al., 2013; Rupp & Leighton, 2016; Smith, 2019). There is a bank of questions that instructors can utilize to compile their assessment, therefore saving them time. Moreover, items designed to be auto-graded usually do not require a high level of reading comprehension (Kan et al., 2019). High-stakes assessments, such as SATs or GCSEs, use E-assessments exclusively. In many cases, students can get immediate feedback. Several features make E-assessment of exams even more pleasing; for example, each student will receive a unique set of numerical values, the order of items will be shifted among students, time limits will be applied per exam or item, submissions can be restricted, and items are displayed one at a time.

2.3. Challenges of E-assessments

Regardless of the benefits of E-assessments, there are some limitations. First, the lack of technical skills, limitations of access, and poor infrastructure might affect the assessment process (Alruwais et al., 2018). In addition, the items' format has a substantial influence (Kan et al., 2019). The use of MCQ is widespread in the E-assessment process in mathematics. Recent research claimed MCQ assessment had been proven effective (Azevedo, 2015) and that students' performance does not differ between an MCQ or long-response format

(Lawrence & Singhania, 2004). However, as Kan et al. (2019) suggested, different item formats require distinct cognitive demands and skills to solve the mathematical problem. It is thus essential to match the item's format to the assessment goal. However, the format of the items in E-assessment is restricted to MCQ, true/false, matching, or short answer submission, so some aspects of students' mathematical understanding—such as their logical and analytical skills—might not be addressed (Wang, 2011). Moreover, such formats might not reflect an accurate evaluation of students' understanding, as some students might simply guess the correct answer. In fact, Stankous (2018) argued that only constructive response items are effective in assessing mathematical competencies.

Mathematics is a subject that requires the knowledge and skills to analyse a problem accurately before solving it. The process (i.e., the sequential steps necessary to solve the mathematical problem) plays a significant role in assessing whether students have achieved the aim of the subject (Rešić & Halilčević, 2014). Evaluating problem-solving problems is thus essential but cannot be accomplished with MCQ or short answer responses (Wang, 2011).

3. Research Methodology

3.1. The WebAssign Platform

WebAssign is an online education platform designed by educators to enrich the teaching and learning experience (<https://webassign.com/>). The students of the Calculus 1 and Business Calculus courses were using the WebAssign platform for all their homework for many years. The platform provides valuable electronic recourses that foster the learning process. For the purpose of assessments, the platform was used only for online homework assignments. However, due to COVID-19 and the lockdown order, the Saudi higher education system converted the educational process into a virtual one. Therefore, the instructors of the two courses chose to conduct each assessment (midterm and final exam) for these two courses on the WebAssign platform. We should mention that the settings for the homework assignments differs from the setting for the exam. For example, when students are doing homework, they are then given up to five attempts to submit answer, in case their previous attempts were wrong, they also have hints and solutions to similar type problem options, as the main purpose of the assignments is to help students learn and master mathematical problem-solving techniques. On the other hand, the setting for the exam was completely different, only one attempt per question was permitted and no help or hint was allowed. The duration of the exam was limited to 60 minutes for midterm and two hours for the final exam.

3.2. Design and Data Collection

The total number of students who started the Calculus I and Business Calculus courses was 62, and the number of students who actually completed both courses was 55. Thus, our sample consisted of a total of 55 freshman and sophomore students registered on the Calculus I and Business Calculus courses (46 students in Calculus I, and 9 students in Business Calculus). Both courses were taught by the authors of this paper during spring 2019/2020. The study

investigated midterm and final exam items and students' responses in these two courses. It is important to mention here that all assessment items were selected from the large questions bank provided by WebAssign. The teachers were not able to develop new items or modify the existing ones. All assessment items were either MCQs or short answer submissions. For the Calculus I course, the midterm had a total of 29 items, the final exam had a total of 36 items, while Business Calculus had a total of 18 items in the midterm exam, and the final exam had a total of 23 items. This study addressed the following three questions:

Q1: Was there a significant difference in students' online midterm and final exam scores between auto-graded and then manually graded scores for the Calculus 1 course?

Q2: Was there a significant difference in students' online midterm and final exam scores between auto-graded and then manually graded scores for the Business Calculus (B.C.) course?

Q3: Were there some common mistakes committed by students during the online exams due to their unfamiliarity with the WebAssign platform, and did this result in zero credit when auto-graded?

3.3. Limitation of the Study

This study considers only two courses, Calculus I and Business Calculus courses taught by both authors at a private university, with a sample size of 55 students.

3.4. Data Analysis

After each online exam, either midterm or final, the exam was immediately auto-graded and scores were recorded, then, manual grading began. Thus, instructors reviewed each student's answers and their auto-grading score, then by using human intelligence it was manually-graded. Thus, depending on the student's mistakes, either full, partial, or no credit was awarded. The manual grading scores were recorded separately for later analysis. The study addressed three questions, and the data analysis is as follows:

(Q1) and (Q2), the quantitative data were analysed using SPSS statistical tools. Thus, using a paired t-test, we explore whether the differences in the auto-graded and manually graded mean scores is statistically significant or not at $\alpha = 0.05$. The paired t-test is robust to non-normality (Schmider et al., 2010).

Regarding the qualitative data in (Q3), while instructors were manually grading the exam they recorded the types of common mistakes committed by students. Those mistakes were characterized into two categories: non-mathematical mistakes (related to WebAssign Platform Design) and minor mathematical mistakes, which might deserve partial credit. All those common mistakes were recorded and samples were collected for the most common types of mistakes. One of the main reasons for this categorization of students' common mistakes is that we can educate our students and try to familiarize them more about the WebAssign platform and the right method to submit their answers so that they take extra care when submitting answers in the future.

4. Results

4.1. Answer to Q1

Q1: Was there a significant difference in students' online midterm and final exam scores between auto-graded and then manually graded scores for the Calculus 1 course?

Students of the Calculus I course were given midterm and final exams on the WebAssign website, using the platform's questions bank. The midterm exam had 29 items (5 MCQs and 24 short answer submissions), while the final exam had a total of 36 items (9 MCQs and 27 short answer submissions).

Each exam, whether a midterm or final was immediately auto-graded, then a manual-grading was carried out. Each item's answer was examined. In particular, when a student obtained zero credit, their answer was analysed. The grades were assigned according to the type of mistakes committed by the student (more detail is provided on this when answering Q3). The differences in the grades between auto-grading and manual-grading were recorded.

The paired t-test revealed a statistically significant difference in students' midterm mean scores at $p = 0.000 < 0.05$. (Table 1). A similar investigation for final exam grades revealed a statistically significant difference in students' mean scores at $p = 0.000 < 0.05$. (Table 2).

Table 1. Significant difference in auto and manual grading for midterm exam of Calculus I.

Calculus 1							
Midterm Exam	Std.		Std. Error		t	df	Sig. (2-tailed)
	Mean	Deviation	Mean				
Auto -Graded	-2.79948	2.08683	.30121	-9.294	45	.000	
Manual-Graded							

Table 2. Significant difference between auto and manual grading for final exam of Calculus I.

Calculus 1							
Final exam	Std.		Std. Error		t	df	Sig. (2-tailed)
	Mean	Deviation	Mean				
Auto - Graded	-1.649565	1.443055	.212767	-7.753	45	.000	
Manual -Graded							

The auto-graded midterm scores for Calculus I had a mean score of 38.04 with 7.351 standard deviation, when it was manually-graded the midterm scores had a mean of 40.84 with 6.45 standard deviation. As for the final exam, the auto-graded students had a mean score of 37.82 with 8.36 standard deviation, and when manually-graded, it had a mean score of 39.47 with 7.74 standard deviation. (Figure 1)

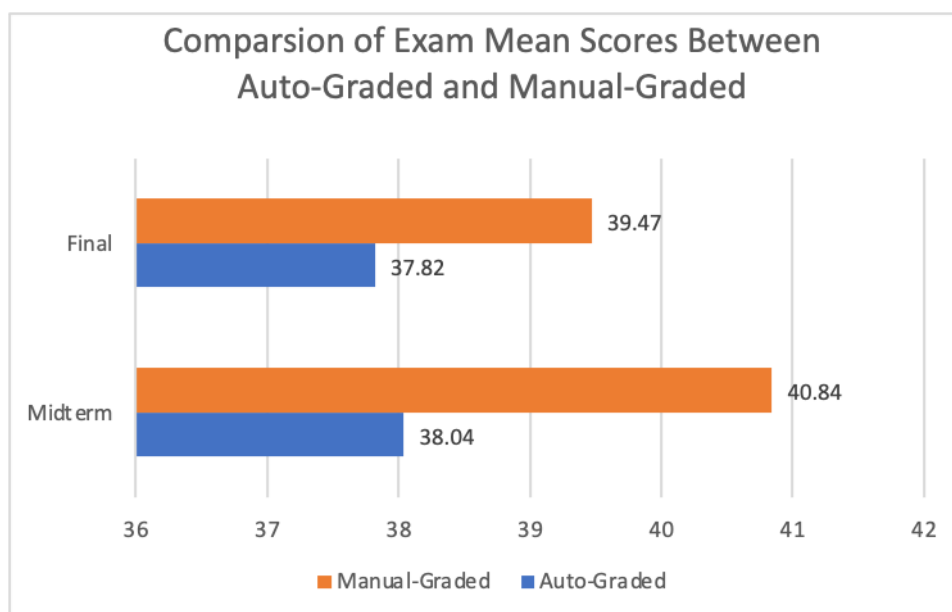


Figure 1. The difference in the mean scores between auto-graded and manual-graded exams for the Calculus I course.

4.2. Answer to Q2

Q2: Was there a significant difference in students' online midterm and final exam scores between auto-graded and then manually graded scores for the Business Calculus (B.C.) course?

Students of the Business Calculus course were given midterm and final exams on the WebAssign website, using the platform's questions bank. They had a total of 18 items in the midterm (3 MCQs and 15 short answer submissions) and the final exam had a total of 23 items (no MCQs, all short answer submission).

Each exam was immediately auto-graded, then a manual-grading was carried out. Each item's answer was examined. In particular, when a student obtained zero credit, their answer was analysed. The grades were assigned according to the type of mistakes committed by the student (more detail is provided on this when answering Q3). The differences in the grades between auto-grading and manual-grading were recorded.

The paired t-test revealed a statistically significant difference in students' midterm mean scores at $p = 0.002 < 0.05$. (Table 3). The final exam results were little different. Observation showed that students committed fewer mistakes on the online final exam than in the midterms. The paired t-test gave a p -value of $p = 0.101 > 0.05$, indicating that the mean differences were not statistically significant (Table 4).

Table 3. Significant difference between auto and manual grading for midterm exam of B.C.

BC Midterm Exam	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-paired)
Auto Graded	-9.72222	6.66667	2.22222	-4.375	8	.002
Manual Graded						

Table 4. The difference between auto and manual grading for the final exam of B.C.

BC Final Exam			Std. Error		t	df	Sig. (2-paired)
	Mean	Std. Deviation	Mean				
Auto -Graded	-2.44667	3.95505	1.31835		-1.856	8	.101
Manual Graded							

The auto-graded midterm grades for the Business Calculus course had a mean score of 20.14 with 12.52 standard deviation, while the manual-graded had a mean score of 29.86 with 11.24 standard deviation. As for the final exam, the auto-graded students had a mean score of 37.302 with 12.4 standard deviation, while the manual-graded students had a mean score of 39.749 with 11.816 standard deviation (Figure 2).

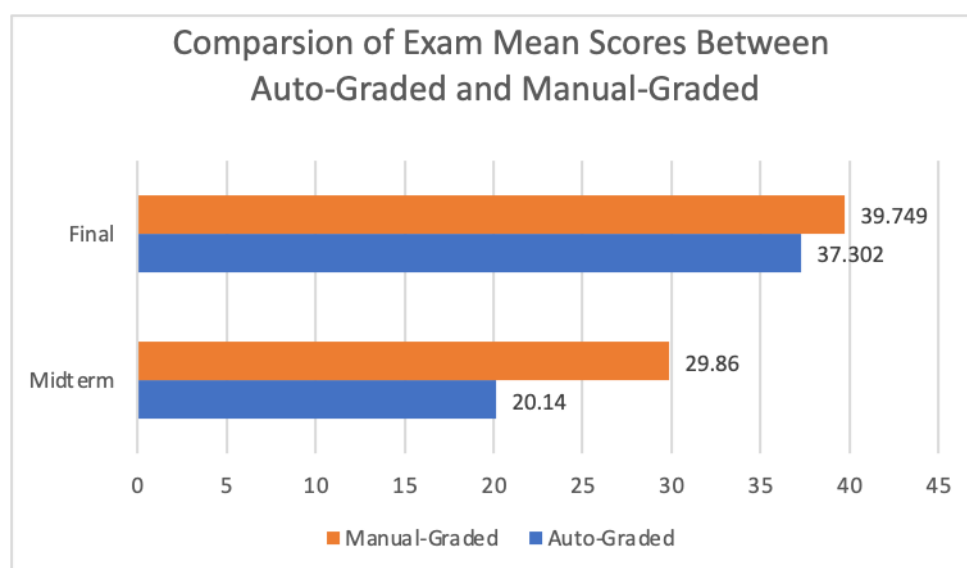


Figure 2. The difference in the mean scores between auto-graded and manual-graded exams for the Business Calculus course.

4.3. Answer to Q3

Q3: Were there some common mistakes committed by students during the online exams due to their unfamiliarity with the WebAssign platform, and did this result in zero credit when auto-graded?

To investigate whether students' unfamiliarity with the platform affected their overall assessment, the authors analysed each student's response to each item on the midterm and final exam. Common mistakes were characterized into two categories:

- 1- Platform Design: In some cases, students did not follow the appropriate website format and hence lost the credit for that item. These are not mathematical mistakes, nor do they represent a lack of comprehension – they only indicate students' unfamiliarity with the website requirements.
- 2- Partial Credits: In other cases, students committed minor mathematical mistakes; they deserved partial credit but received no credit.

4.3.1. Platform Design

Some students did not adhere to the typical writing design required by the software platform of the WebAssign. They used different symbols, capitalization, or representations.

The following samples represent innocent "mistakes" committed by students while submitting their answer on the platform, it does not indicate a lack of students' mathematical skills or knowledge:

- 1- When solving an indefinite integral task, a student performed the integration accurately but wrote the integration constant as a small letter 'c' instead of a capital letter 'C', and the constant of integration is just a symbol, thus writing it as a small or large letter does not alter the answer, but this mistake resulted in zero credit (Figure 3).
- 2- Instead of writing the variable "x," some students used the Greek letter chi " χ ", which looks similar to the "x," and all computations were correct. This is not a mathematical mistake, but it resulted in zero credit (Figure 4).
- 3- A student computed the derivative of the function correctly. However, when submitting the answer, the student wrote the symbol of the derivative \dot{y} and then the equals sign as usually one does when they write the answer on paper. But on the WebAssign platform, they are supposed to only enter the answer without \dot{y} . This made WebAssign not recognize the correct answer and resulted in zero credit (Figure 5).
- 4- After a long process of taking the derivative and finding the inflection point, the student submitted a correct answer as an ordered pair, and while submitting she inserted brackets as is usually done when we write on paper, she did not notice that the brackets were already included. This made WebAssign not recognize her correct answer due to extra brackets and she received zero credit. (Figure 6).
- 5- The definite integral computation was all correct. The students submitted the answer in a decimal format instead of a fraction. This resulted in zero credit (Figure 7).
- 6- This is a Business Calculus problem; it is a long problem related to finding the producer's surplus. When submitting the final answer, the student rounded it to one decimal place instead of two decimal places (89.3 instead of 89.33). All computations were accurate, and this resulted in zero credit (Figure 8).

Samples of students who submitted a correct answer and received zero credit when auto-graded.

0/2 points | Previous Answers HarMathAp12 12.3.005. [4025659]

Evaluate the integral. (Use C for the constant of integration.)

$$\int 3000e^{0.1x} dx$$

Remember to use capital C.

Figure 3. Using the lowercase letter "c" as the integration constant

3. 0/2 points | Previous Answers HarMathAp12 11.2.011. [4024349]

Find the derivative of the function.

$$y = 2e^{(x^2 + 6)^3}$$

Figure 4. Using the Greek letter chi "χ" instead of "x"

5. 0/2 points | Previous Answers HarMathAp12 11.1.009. [4024390]

Find the derivative of the function.

$$y = \ln(5x^2 - x) + 8x$$

Figure 5. Repeated article Ψ

(c) Find the inflection points.

$(x, y) = ($ $)$ (smaller x-value)

$(x, y) = ($ $)$ (larger x-value)

Figure 6. Extra brackets

16. 0/1 points | Previous Answers HarMathAp12 13.2.051. [4025575]

In this problem, use properties of definite integrals.

If $\int_1^2 (4x - x^2) dx = \frac{11}{3}$ and $\int_2^5 (4x - x^2) dx = 3$, what does $\int_1^5 (x^2 - 4x) dx$ equal? (Give an exact answer. Do not round.)

Figure 7. Using decimal format

0/2 points | Previous Answers HarMathAp12 13.4.027. [4025845]

In this problem, p is in dollars and x is the number of units.

Suppose that the supply function for a good is $p = 4x^2 + 34x + 2$. If the equilibrium price is \$86 per unit, what is the producer's surplus there? (Round your answer to the nearest cent.)

\$89.3 89.33

Figure 8. Rounding to one decimal place instead of two decimal places

4.3.2. Partial Credits

Sometimes students commit minor mathematical mistakes, which result in partial credit when the problem is manually graded. Unfortunately, the current auto-grading system does not offer partial credit. Here is a list of examples of minor mathematical mistakes which deserve partial credit.

- 1- The student computed the derivative but while entering the answer the student missed the negative (-) sign in one place. All other computations were accurate. When we manually grade such a problem the student deserves partial credit, but auto-grading resulted in zero credit. (Figure 9).
- 2- The student found the derivative correctly but when submitting her answer on the platform, in one place she missed a single variable ``x``, which she forgot to write. However, the WebAssign platform cannot distinguish this small error, so it resulted in zero credit, however when such a problem is graded manually then she absolutely deserves partial credit (Figure 10).
- 3- The student found the derivative correctly but when submitting her answer on the platform she misplaced one single bracket in a multi-bracket item, this resulted in zero credit when auto-graded (Figure 11).
- 4- The student found the derivative correctly but when submitting her answer on the platform, in one place instead of writing variable ``t``, by mistake she wrote variable "x". This is just an answer submission mistake, the auto-grading system does not distinguish this small mistake and results in zero credit (Figure 12).

3. 0/2 points | Previous Answers SCalcET8M 3.2.021. [3458628]

Differentiate.

$$f(t) = \frac{\sqrt[3]{t}}{t-3}$$

$$f'(t) = \frac{2t+3}{3\sqrt[3]{t^2}(t-3)^2} \quad \times \quad \frac{-2t-3}{3(t-3)^2t^{2/3}}$$

Need Help?

Figure 9. Missing a negative sign (-)

2. + 0/2 points | Previous Answers HarMathAp12 9.5.020. [4025200]

Find the indicated derivative.

$$f'(x) \text{ for } f(x) = \frac{(x+3)(x-4)}{x^2+6}$$

$$f'(x) = \frac{x^2+36-6}{(x^2+6)^2} \quad \times \quad \frac{x^2+36x-6}{(x^2+6)^2}$$

Need Help? Read It Watch It Talk to a Tutor

Figure 10. Missing a variable "x" in a single place

+ 0/2 points | Previous Answers SCalcET8M 3.4.031. [3458615]

Find the derivative of the function.

$$F(t) = e^{2t \sin(2t)}$$

$$F'(t) = e^{2t \sin(2t)} (2 \sin(2t)) + 4t \cos(2t) \quad \times \quad e^{2t \sin(2t)} (4t \cos(2t) + 2 \sin(2t))$$

Figure 11. Misplacing a bracket

4. + 0/1 points | Previous Answers SCalcET8M 3.4.031. [3458615]

Find the derivative of the function.

$$F(t) = e^{9t \sin(2t)}$$

$$F'(t) = e^{9t \sin(2t)} [9 \sin(2t) + 18t \cos(2x)] \quad \times \quad e^{9t \sin(2t)} (18t \cos(2t) + 9 \sin(2t))$$

Figure 12. Using the variable "x" instead of "t" in single place by mistake

5. Discussion

This study aimed to evaluate the E-assessment process used in two calculus courses at one higher education institution in the KSA. Students were used to submitting online homework on the WebAssign platform eight weeks before the lockdown. However, during exams, it is understandable that students might commit the aforementioned non-mathematical mistakes while submitting their answers due to exam pressure, the limited time duration, and the restriction of a single submission attempt. It must be noted that mistakes provoked by the platform's design are only relevant during an E-assessment process. Unfortunately, WebAssign's auto-grading system will either assign a full credit (100%) or no credit (0%) to an item. No partial credit is considered. This issue is consistent with the limitations reported by Alruwais et al. (2018), Kan et al. (2019) and Smith (2019).

The statistical analysis results demonstrated a significant difference between auto-graded and manual-graded exams scores for both midterm and final exams for the Calculus I course (Tables 1 and 2). Whereas the bar chart (Figure 1) displayed the differences in the mean scores between auto-graded and manual-

graded midterm and final exams for the Calculus I course. The literature indicated a discrepancy between auto-grading and manual-grading (Bejar et al., 2017), which is consistent with our research results.

As for Business Calculus, the result was a little different. Table 3 displayed statistically significant differences in the midterm mean scores between the auto-graded and manually-graded exams. However, for the final exam, the difference was not statistically significant (Table 4). One factor that must be considered is that the Business Calculus course is mandatory for the Finance programme only; hence it has a lower number of students than Calculus I (which is offered for Computer Information, Engineering, and Communication & Network programmes). In the online midterm exam of Business Calculus, most students' mistakes were related to platform design rather than mathematical mistakes. Thus, they became meticulous about submitting their answers on the WebAssign platform by growing their cognizance of the WebAssign platform's tools and continually asking the instructor for every detail to minimize their mistakes related to the platform design. The final exam result demonstrated fewer mistakes related to platform design. This revealed the importance of educating our students in detail about online platform design. The bar chart (Figure 2) displayed the differences in the mean scores between auto-graded and manual-graded midterm and final exams for the Business Calculus course.

To investigate whether students' unfamiliarity with the platform affected their overall assessment, the authors analysed each student's response to each item on the midterm and final exam. The common mistakes were characterized into two categories: Platform Design and Partial Credits. By examining the samples of answers provided by students, it becomes clear that students deserved either partial or full credit when auto-graded awarded them zero credit.

The assessment items' type influences the assessment's validity. The results reflected a surprising number of inconsistencies between the auto-grading and manual-grading. Figures 3-8 illustrated the most common mistakes committed by students that resulted in zero credit when auto-graded.

In many cases, the mistake was only due to a lack of familiarity with the formatting, not a lack of mathematical understanding (Figures 3-8). For example, submitting answers with additional brackets (Figure 6) resulted in zero credit for the answer, this mistake was apparent when students wrote an ordered-pair number (an inflection point), intervals of an increasing and decreasing function, or a function's concavity status. This is consistent with what Smith (2019) stated – that formatting and technical issues could result in losing credits.

In some other cases, if the student made a minor mathematical mistake, then she/he deserves a partial credit, which is a standard agreement in mathematics assessments. For example, after carrying out a complicated derivative task, a student incorrectly missed a "minus sign" (Figure 9). In the auto-grading process, the student received zero credit for the item, whereas if human intelligence were applied then partial credit would have been awarded.

As our analysis demonstrated, E-assessment has some limitations. It has been argued that students might misinterpret an item or some minor terminology and hence fail to choose the correct answer (Wang, 1998). Moreover, sometimes students commit simple mistakes if they misread an item in E-assessment, which can otherwise be addressed with short feedback from the instructor. However, this is not possible with auto-grading (Smith, 2019), which is consistent with our research results.

Moreover, considering partial grading is essential. Even if a student must only submit the final answer, this final answer can be evaluated from several angles. In some cases, students do not submit a 100% accurate answer, but a partially valid answer deserves partial credit. Unfortunately, this is not always possible with auto-grading (Smith, 2019), as our research indicated.

6. Conclusion

The demand for E-learning and E-assessments are increasing, especially during the recent COVID-19 pandemic as traditional teaching and assessments methods are no longer favourable. In this study, we have investigated the validity and appropriateness of E-assessments in Calculus courses, by comparing the difference in the means scores of auto-graded and manually graded exam scores. The differences were statistically significant, thus illustrating the importance of manual grading and its role in assessing students' achievements and mathematics understanding. From the results, we can conclude that we cannot solely rely on auto-grading, as it will not be a valid indicator of a student's mathematical proficiency. Also, a partial score is essential; however, this aspect is not considered in auto-grading. Hence, auto-grading is a useful approach, but it has some limitations and needs to be more intelligent. E-learning and E-assessments put an extra non-Mathematical burden on the students – the need to learn about each online platform design, which they will use. Therefore, students need to master new skills (when using a computer to submit answers), and a lack thereof will affect their success (Parshall & Guille, 2015). Our study has characterized the most common type of mistakes committed by students during online exams. We hope our research will guide the software designer to target those common mistakes and upgrade their auto-grading software to account for partial credit and to correct platform design errors.

Acknowledgment

The authors would like to thank the Deanship of Educational Services at Prince Sultan University for their support. Also, the authors would like to thank the reviewers for their helpful comments, which helped the paper takes its current shape.

Declaration

The authors indicate no conflict of interest.

9. References

- Abbasi, S., Ayoob, T., Malik, A., & Memon, S. I. (2020). Perceptions of students regarding E-learning during COVID-19 at a private medical college. *Pakistan Journal of Medical Sciences*, 36, 57–61. <https://doi.org/10.12669/pjms.36.COVID19-S4.2766>

- Alanazi, A. A., & Alshaalan, Z. M. (2020). Views of faculty members on the use of e-learning in Saudi medical and health colleges during COVID-19 pandemic. *Journal of Nature and Science of Medicine*, 3(4), 308-317. https://doi.org/10.4103/JNSM.JNSM_82_20
- Alqahtani, A. Y., & Rajkhan, A. A. (2020). E-Learning critical success factors during the COVID-19 pandemic: A comprehensive analysis of E-Learning managerial perspectives. *Education Sciences*, 10(9), 1-16. <https://doi.org/10.3390/educsci10090216>
- Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, 8(1), 34-37. <https://doi.org/10.18178/ijiet.2018.8.1.1008>
- Azevedo, J. M. (2015). E-Assessment in mathematics courses with multiple-choice questions tests. In *Proceedings of the 7th International Conference on Computer Supported Education, CSEDU-2015* (pp. 260-266). <https://doi.org/10.5220/0005452702600266>
- Baki, A., Kosa, T., & Guven, B. (2011). A comparative study of the effects of using dynamic geometry software and physical manipulative on the spatial visualization skills of pre-service mathematics teachers. *British Journal of Educational Research*, 42, 291-319. <https://doi.org/10.1111/j.1467-8535.2009.01012.x>
- Bejar, I. I., Mislevy, R. I., & Zhang, M. (2017). Automated scoring with validity in mind. In A. A. Rubb, & J. P. Leighton, *The handbook of cognition and assessment: Frameworks, methodologies and applications* (pp. 226-247). Wiley-Blackwell.
- Broughton, S. J., Robinson, C. L., & Hernandez-Martinez, P. (2013). Lecturers' perspectives on the use of a mathematics-based computer-aided assessment system. *Teaching Mathematics and its Application*, 32, 88-94. <https://doi.org/10.1093/teamat/hrt008>
- Camacho Machín, M., Depool Rivero, R., & Santos-Trigo, M. (2010). Students' use of Derive software in comprehending and making sense of definite integral and area concepts. *CBMS Issues in Mathematics Education*, 16, 29-61.
- Fulton, C. (2020). Collaborating in online teaching: inviting e-guests to facilitate learning in the digital environment. *Information Learning Sciences*, 121(7/8), 579-585.
- Heid, M. K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education*, 19, 3-25.
- Hodges, C., Moore, S., Lockee, B., Trust, T. & Bond, M. A. (2020, March 27). The difference between emergency remote teaching and online learning. *Educause Review*. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning#fn12>
- Kan, A., Bulut, O., & Cormier, D. C. (2019). The impact of item stem format on the dimensional structure of mathematics assessments. *Educational Assessment*, 24(1), 13-32. <https://doi.org/10.1080/10627197.2018.1545569>
- Kerzic, D., Umek, I., Tomazevic, N., & Aristovnik, A. (2018). Evaluating the impact of e-learning on students' perception of acquired competencies in a university blended learning environment. *Journal of e-learning and Knowledge Society*. 14(3), 65-76.
- Khalil, R., Mansour, A. E., & Fadda, W. A. et al. (2020). The sudden transition to synchronized online learning during the COVID-19 pandemic in Saudi Arabia: a qualitative study exploring medical students' perspectives. *BMC Medical Education*, 20(285), 1-10. <https://doi.org/10.1186/s12909-020-02208-z>
- Lawrence, J. A., & Singhanian, R. P. (2004). A study of teaching and testing strategies for a required statistics course for undergraduate business students. *Journal of*

- Education for Business*, 79(6), 333–383. <https://doi.org/10.3200/JOEB.79.6.333-338>
- McCarthy, K. (2020). "The Global Impact of Coronavirus on Education." [Video]. *ABC News Network*. <https://abcnews.go.com/International/global-impact-coronavirus-education/story?id=69411738>
- Parshall, C. G., & Guille, R. A. (2015). Managing ongoing changes to the test: Agile strategies for continuous innovations. In F. Drasgow (Ed.), *Technology and Testing: Improving Educational and Psychological Measurement*. National Council of Teachers of Mathematics.
- Pennisi, E. (2020). Courses bring field sites and labs to the small screen. *Science*, 369(6501), 239–240. <https://doi.org/10.1126/science.369.6501.239>
- Raja, R., & Nagasubramani, P. C. (2018). Impact of modern technology in education. *Journal of Applied and Advanced Research*, 3(1), 33–35. <https://doi.org/10.21839/jaar.2018.v3S1.165>
- Rešić, S. & Halilčević, S. (2014). Tracking, evaluation, and grading students' accomplishments in mathematics classes. *Human: Journal for Interdisciplinary Studies*, 4(3), 41–50.
- Rešić, S., Bajramović, A., & Hrnjičić, A. (2017). Grade as the motivational factor in learning mathematics. *Human: Journal for Interdisciplinary Studies*, 7(2), 10–21. <https://doi.org/10.21554/hrr.091702>
- Rupp, A. A., & Leighton, J. P. (2017). Introduction to handbook. In A. A. Rubb, & J. P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies and applications* (pp. 1–11). Wiley-Blackwell.
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6, 147–151.
- Shahbari, J. A., & Abu-Alhija, F. N. (2018). Does training in alternative assessment matter? The case of prospective and practicing mathematics teachers' attitudes toward alternative assessment and their beliefs about the nature of mathematics. *International Journal of Science and Mathematics Education*, 16, 1315–1335. <https://doi.org/10.1007/s10763-017-9830-6>
- Smith, J. S. (2019, June), *Determining optimal deployment strategies of MATLAB autograder to maximize student learning and engagement*. Paper presented at 2019 ASEE Annual Conference & Exposition, Tampa, Florida. <https://peer.asee.org/32627>
- Stankous, N. V. (2018). Constructive response vs. Multiple-Choice tests in math: American experience and discussion (Review). *European Scientific Journal* (special ed.), 308–316. <https://core.ac.uk/download/pdf/328025438.pdf>
- Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *Lancet Child & Adolescent Health*, 4, 397–404. [https://doi.org/10.1016/S2352-4642\(20\)30095-X](https://doi.org/10.1016/S2352-4642(20)30095-X)
- Wang, J. (1998). A content examination of the TIMSS results. *Phi Delta Kappan*, 80(1), 36–38.
- Wang, J. (2011). Re-examining test item issues in the TIMSS Mathematics and Science Assessments. *School Science and Mathematics*, 111(7), 33–34.