Comparison of Exam Scores and Time Taken on Exams Between Proctored On-Campus and Unproctored Online Students

Daris Howard
Brigham Young University–Idaho

Abstract
Online education is expected to grow, bringing new challenges. One of the biggest challenges concerns the validity of online assessments. Questions arise about cheating, including whether or not the person taking the assessments is the student registered for the course. Studies have tried to determine the amount of cheating in online assessments using student self-reporting. Concern about the validity of these studies has led to quantitative studies attempting to determine the level of cheating in online classes by comparing unproctored online classes against proctored classes. This quantitative quasi-experimental study uses such an approach, comparing test scores and the amount of time online unproctored students spend on exams against test scores and the amount of time spent on exams by students proctored in a testing center and by students proctored online using software. The data for each of the three groups, online-unproctored, testing-center-proctored, and software-proctored, were collected for each of the three unit exams, with data samples over 1,000 for each group on each test. The means of the exam scores of the unproctored online students were similar to the means of the exam scores of the other two groups, but the means for the amount of time the unproctored online students spent on the exams were significantly greater than were the means of the time spent by students in the either of the other two groups. The increased amount of time spent by the unproctored students likely indicated that students looked up answers during the tests, even though looking up answers was prohibited.

Keywords: online education, academic integrity, online testing, proctoring software, testing-center proctoring, cheating, testing validation, assessment

Howard, D. (2020). Comparison of exam scores and time taken on exams between proctored on-campus and unproctored online students. *Online Learning, 24*(4), 204-228. https://doi.org/10.24059/olj.v24i4.2148

Comparison of Exam Scores and Time Taken on Exams Between Proctored On-Campus and Unproctored Online Students

The number of students in online education is expected to rise for some time (National Center for Educational Statistics [NCES], 2016). The growth of online education creates issues that are increasingly important to higher education. One of the most important issues deals with the integrity of online work, especially with possible cheating in online testing (Berkey & Halfond, 2015; Corrigan-Gibbs, Gupta, Northcutt, Cutrell, & Thies, 2015; Madara & Namango, 2016).

This work, which comes from a dissertation (Howard, 2019), considered the problems with possible cheating on unproctored online exams in a general education (GE) math course to determine the scope of the problem without using student self-reporting. Many past studies have used unreliable student self-reporting in qualitative studies to examine the problem (Brown, 2018). A few studies have used quantitative approaches, with some using online test proctoring software (Alessio, Malay, Maurer, Bailer, & Rubin, 2017; Daffin & Jones, 2018).

This study considers whether there are significant differences in GE math exam scores and the amount of time taken on the exams when online students are unproctored, in comparison to on-campus students who are proctored. Increased exam scores or increased time on exams could be an indication of cheating on unproctored online exams. The comparison of unproctored exams against physically proctored and software proctored exams can also help determine if the proctoring method makes a difference.

Review of Relevant Literature

Many studies considered cheating in a qualitative way, asking the students if they cheated or had seen cheating in others (Brown, 2018). The question arises as to whether students who would cheat in assessment could be relied on to give honest answers in a qualitative study. For that reason, this research attempted to study the concept of cheating quantitatively without the need for relying on the students' answers.

The research on the subject was also somewhat contradictory and inconclusive. Some claimed that students cheat less online than in a face-to-face class (Ladyshewsky, 2015), while others indicate there is significantly more cheating in unproctored classes (Alessio, Malay, Maurer, Bailer, & Rubin, 2017; Daffin & Jones, 2018).

Some early research that attempted to do quantitative studies did so by comparing scores of students tested online without proctoring to students tested in physically-proctored testing centers. In all of these studies except one, the unproctored students scored higher. The one exception where the proctored and unproctored students scored basically the same stood out due to the fact that the test was based on skills in Word, PowerPoint, Excel, and Access (Hollister & Berenson, 2009).

One important note is that the only studies that suggested there was less cheating when a student was not proctored were the qualitative ones that relied on the student responses to surveys. All quantitative studies except the one by Hollister and Berenson (2009) indicated there was likely more cheating in online unproctored classes.

However, the qualitative studies did give some interesting information. Some key points included the fact there appears to be an increase in cheating in general, with as high as 81% to 85%

of students admitting to cheating (Polat, 2017). Another big concern was proxy or impersonation cheating, which is when a student hires someone else to take an assessment, or even the whole class, in the place of the student. In a study in the Philippines, proxy cheating was the number one problem students mentioned that they had seen concerning cheating, with 72% of the students saying they knew students (themselves or others) who had engaged in this practice (Ravasco, 2012). Some companies blatantly offer services in this area, offering everything from prewritten research papers to having someone take the full course (Watson & Sottile, 2010).

Another important note that came from the research was the law regarding the verification of students in assessment. The Higher Education Act (110th U.S. Congress, 2008) says,

> (ii) the agency or association requires an institution that offers distance education or correspondence education to have processes through which the institution establishes that the student who registers in distance education or correspondence education course or program is the same student who participates in and completes the program and receives the academic credit (p. 249).

This law says, among other things, that an institution of higher education must be able to verify the identity of those taking the course. The law was passed twelve years ago, and still many, if not most, higher educational institutions are not entirely in compliance. Accrediting bodies, which had formerly been mostly involved in determining the value of the curriculum, are now starting to make concerns over cheating, especially impersonation cheating, an issue (Bemmel, 2014).

With these ideas in mind, the goal of this study was to formulate a method to determine problems with cheating without using questionnaires. The other quantitative studies had used scores and the determination was made to use that as one piece. The other piece would be a comparison of the amount of time the students spent on the test. Another aspect that was deemed important was to use modern technology, such as using software proctoring. During the time this study was being put together, two studies with some similar research were released. The results of these two studies are beneficial to consider before the results of this study are presented.

The first study dealt with an elective course in which the students learned medical terms (Alessio et al., 2017). There were nine sections, and four chose to use software proctoring. Using proctoring was voluntary on the part of the instructor, and the instructor determined which tests to proctor, so the findings were not completely random. The students did not know until starting the test whether they would be proctored. The researchers created a formula in an attempt to equalize the proctoring. The results showed that the mean test score for proctored students was 74.3% (SD = 12.3), and the mean test score for unproctored students was 89.4% (SD = 9.0). This a difference of a C to an A- in grades from proctoring to unproctored, and the unproctored student used more than twice as much time taking the test on average. The amount of time between the two groups was also significant, with unproctored students averaging 20.4% (SD = 13.9) of the allotted time, and proctored students using an average of 41.2% (SD = 14.1) of the allotted time.

The second study had similar results (Daffin & Jones, 2018), with a somewhat different design. In the Daffin and Jones study, both physical and software proctoring were compared against unproctored testing. The study was done in psychology classes from entry-level to graduate level. The students knew ahead when they would be proctored. The results came from a student's paired scores, not scores compared with different students. Each student took one proctored test, and the others were not proctored. Not all students were proctored on the same tests in order to account for differences in the tests. When proctored, the student test averages were 64.37% (SD = 15.47), with a time of 27.69 (SD = 5.17) minutes. The unproctored students had a mean on the test scores of 77.12% (SD = 15.47), with the average time being 48.18 (SD = 5.13) minutes. This is a difference of a D to a C+, with 74% more time taken on unproctored exams.

These two recent studies mentioned give a strong indication that students who were unproctored were looking up material (increase in time) and were likely cheating (Alessio et al., 2017; Daffin & Jones, 2018). This research was to have a similar approach with a few differences. The class that would be evaluated was Math for the Real World, a required GE math class, one which many students may reason is not essential to their degree.

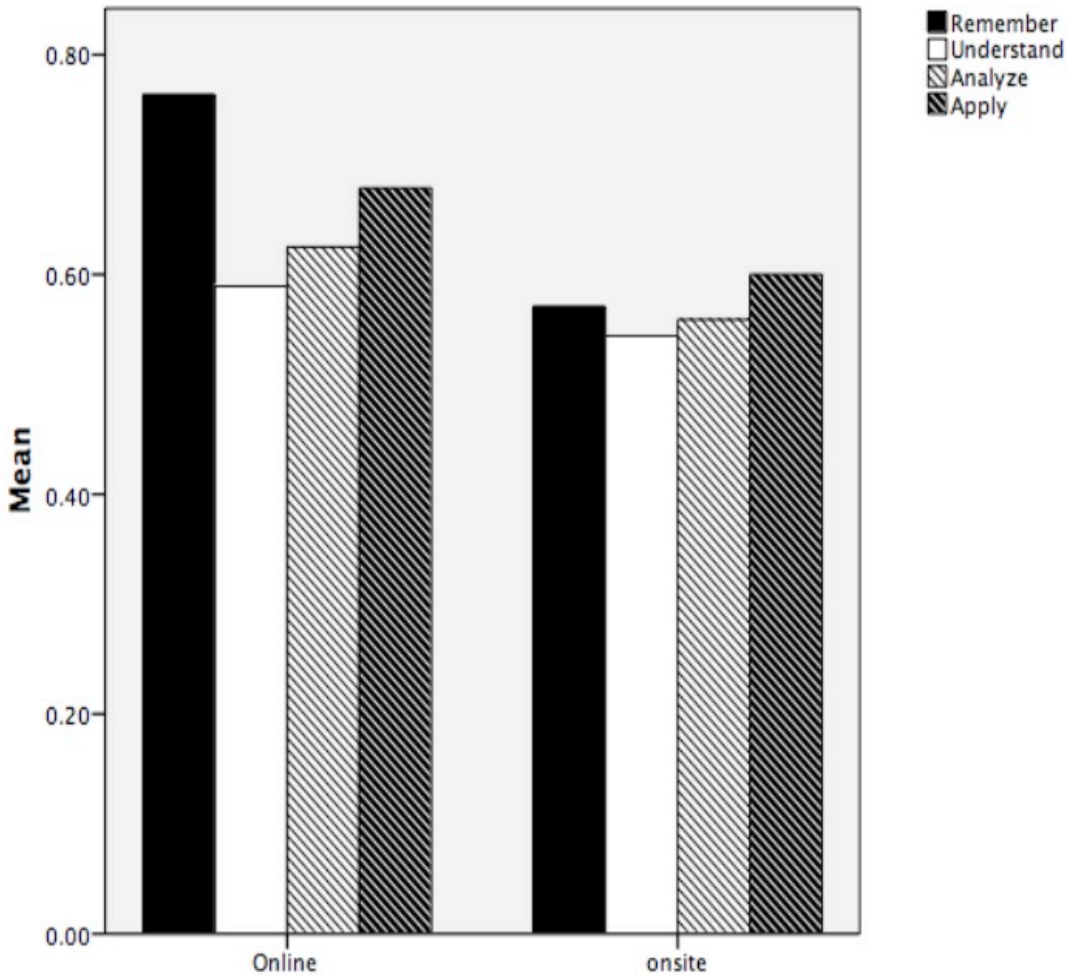
There are some other studies from the literature also worth mentioning. In a study by Richardson and North (2013), the researchers point out two additional, interesting pieces of information. The first was once a class was proctored, the exam averages tended to be lower for subsequent unproctored exams for the class when compared to classes which were not previously proctored. The reasoning given was that once students' courses were proctored, the proctoring seemed to linger in the students' minds, reducing cheating even when the student was not proctored. The second piece of information pointed out was the difference between when the course was undergraduate compared to graduate-level courses. The graduate classes showed less change in exam scores based on proctoring than the undergraduate classes did. Richardson and North postulated that the more mature a student is, the less likely the student is to cheat. This analysis would match what others have said about maturity reducing cheating (Ladyshewsky, 2015).

Another study that had significance in this study is a comparison of online and on-campus sections (Varble, 2014). In the research by Varble, there was not only a comparison between online and face-to-face teaching, but some other information related to Varble's research stands out, which could correlate to this study. Varble attempted to isolate the teaching method by having two sections, one online and one face-to-face, taught by the same teacher and containing the same material. The online students took all exams online. The face-to-face students took all exams except three (the last two and the final) online, and these last three exams were taken with paper and pencil in the classroom with physical proctoring. The online exams were not proctored, though best practices were used for the online testing, such as randomizing question order, randomizing questions from a test bank, providing only one question at a time without being able to go back, and lockdown of the browser so students could not go to other windows.

In all cases except one in the research by Varble (2014), where both groups of students took the online exams, the online students scored higher. The reason the online student scored higher on most exams might be explained by the fact the online students were more familiar with the online environment. However, even though the online students scored higher when the exams were taken online, the difference was not significant between the two groups. On the other hand, on the three exams where the face-to-face students took the exam in a physically-proctored classroom, the exams showed a significant difference between the two groups, with the face-to-face, proctored students scoring lower. Putting this information together, the indication would be the delivery method was not the issue which created the significant grade difference, but the issue which caused the difference was whether or not the exam was proctored.

One other important point comes from the study by Varble (2014). Varble breaks down the concepts on the final into four categories: (a) *remember*, (b) *understand*, (c) *analyze*, and (d) *apply*. On the final, the category with the biggest change when comparing on-campus proctored students and unproctored online students was the *remember* category. Varble's graph related to the four categories and the online and onsite groups is shown in Figure 1.

Figure 1. Delivery comparison. This figure illustrates the comparison of the categories of final examination items for onsite and online students. Reprinted with permission (Varble, 2014).



Other literature also indicates that the test design can be an essential factor in helping to reduce cheating (Egan, 2018). The purpose of this research is to consider whether there are differences between test scores and the amount of time taken on tests when comparing proctored and unproctored students. Though test design is not the main focus of this research, the tie-in to test design is important when considering the possible options students may have for looking up answers on unproctored tests when it is not allowed.

Methods

The data for this study comes from the GE math course at Brigham Young University – Idaho (BYUI). The on-campus testing-center proctored students' data were from Spring 2018, and the on-campus software-proctored students' data were from Fall 2018. The online unproctored students' data were from Fall 2018. Each data set was greater than a thousand students, providing a large dataset.

The GE math class is designed as a freshmen level class and is usually mostly taken by freshmen in both online and on-campus sections. The comparison for the three groups for the percent of freshmen and non-freshmen, along with the average age of the freshmen and average age of all students in the groups, can be seen in Table 1.

Table 1

Freshmen vs. Non-freshmen in Groups and Age Comparison

Group testing type	Percent of freshmen	Percent of non-freshmen	Average age of freshmen	Average age of all students in groups
On-campus testing-center proctored Spring 2018	69.2	30.8	20	21
On-campus software-proctored Fall 2018	72.2	27.8	19	20
Online unproctored Fall 2018	52.4	47.6	35	37

Data provided by the director of data and analysis at BYUI (Sleight, 2020).

Though the demographics are not the emphasis of this study, it is important to note that there are two big differences between the online and on-campus classes. Even though there are more than 50% freshmen in each group, the percent of freshmen in the online course is significantly lower. Studies indicate that graduate students likely cheat less than undergraduate students (Richardson & North, 2013). The demographic of grade (freshman, sophomore, junior, senior) could be an essential difference for further research.

The second difference is age. The average age of the freshmen in the online class is fifteen to sixteen years older than the average age of the freshmen in the online group. The average age of all students is sixteen to seventeen years older in the online group. The age difference has been shown as a big factor in the possible cheating of students (Ladyshewsky, 2015). Both age and class (freshman, sophomore, junior, senior) are big differences and are something that deserves further research beyond the results shared in this study.

Another consideration about those in this study is the fact that the students who take the GE math course are those who do not fill the math GE requirement in another way. Students in fields requiring calculus or higher statistics, or students transferring to BYUI with GE requirements fulfilled, do not have to take this GE math course. However, most students on campus do end up taking the course. The limiting of those taking the course would be another area of interest for further research.

The prerequisites for the course were the same for both online and on-campus students. A student needed to have an ACT Math score of at least 18, an SAT math score of 430, or an ALEKS test score of 38. A student could also take one of three algebra classes and get a B or better. These classes were offered both on-campus and online.

For this course, the students take three unit exams. The data comparisons were made between the two proctored assessment methods against the unproctored assessments separately for each of the unit exams. There is also a comprehensive final, but since the lowest test was dropped, many students opted out of taking the final, so it was not used in the analysis. The time each student took on each exam was also recorded. Both exam scores and time taken on exams for the two proctored groups were compared against the online unproctored group.

The sample selection for unproctored online classes was all sections for the Fall 2018 semester. No sections were excluded. The sample for the on-campus physically proctored class was all Spring 2018 sections, with no sections excluded. In the data set for the software-proctored on-campus classes, a teacher could receive permission to physically proctor the teacher's own students and not use software proctoring. Two teachers chose this option and were excluded from the software-proctoring set. All the data set samples were of sufficient size to meet the data sample size requirements as defined by Sullivan (Sullivan, 2013).

Some on-campus students take the online class. When registration opened for the GE math class at the university, the students could choose either online or on-campus classes. The on-campus classes filled first, and then the students who were late registering were often forced to take the online classes. Students who registered late may have done so for many reasons, including the students disliked the class, the students lacked skills in the subject, or the students had procrastination tendencies. The registration process created two possible problems: (a) self-selection bias, and (b) inequality between the on-campus and online classes. Both forced selection into classes and self-selecting could create bias (Fendler, Ruff, & Shrikhande, 2018), but registration bias is often part of many educational studies (Stack, 2015). The forced registration into online classes was somewhat offset as students added and dropped classes when the semester started. On-campus, online students could often obtain a seat in an on-campus class if desired.

The number of sections in each group and the number of students who took each exam can be seen in Table 2. The numbers represent the number of students taking the exams, with the number n for the sample size for each group in the analysis. These numbers are the group sizes after the unusable data were discarded. The large datasets help ensure the soundness of the statistical analysis.

Table 2

Sizes of Groups in the Study

Testing type	No. sections	No. original students	No. taking exam 1	No. taking exam 2	No. taking exam 3
Testing center	46	1,211	1,171	1,144	1,092
Software proctored	53*	2,244	1,888	1,869	1,779
Online unproctored	41	1,893	1,691	1,663	1,565

*There were 55 on-campus sections for the Fall 2018 semester, but only 53 were proctored using software. The other two were proctored by the individual teachers and were not included in this study.

The exams in the testing center were all the same. For both the on-campus software-proctored and the online unproctored students, the exam questions were similar to the testing-center exam questions. However, the numbers in the questions were generated by the testing software. The online tests for both the online unproctored students and the on-campus software proctored students were taken in Maple TA (now called Mobius) with access through the Learning Management System (LMS). All exams were multiple choice. The online exams, both software proctored and unproctored, presented only one question at a time and were time-limited to two hours, in line with best practices for online testing (Varble, 2014). There was no time limit on the tests in the testing center except for the limit of the hours the testing center was open.

The data were collected for the testing-center proctored exams in the spring semester of 2018, and the data were collected for both online groups fall semester of 2018. For consistency, the tests were very similar between the two semesters. Keeping the tests similar should have alleviated test questions playing into any difference between the proctoring methods.

A few methods on the data cleaning are worth noting. The data for the online tests, both the online students and campus software-proctored students, showed test times as high as 53 days. The high amount of time meant something was wrong since test times were limited to two hours. In checking further, Maple TA recorded the opening time of the test the first time the test was opened. Maple TA recorded the closing time of the test for the last time the test was taken. What that meant was if a student's internet connection was interrupted, or something else happened, and the teacher opened the test again, the closing time for the test would be that of the last attempt, while the opening time would be of the first attempt. Recording test times this way meant the times on the online tests that were over two hours were invalid since those times were across unused time between multiple exam attempts. Therefore, test times over two hours had to be removed from the data set. Table 3 shows the percentage of tests that were over two hours for each group and each test.

Table 3

Number of Student Exams Over the Two-Hour Limit and the Percentage

Testing type	Exam 1	Exam 2	Exam 3
Testing center	4/1171 0.34%	6/1144 0.524%	6/1092 0.55%
Software proctored	51/1836 2.78%	27/1841 1.47%	20/1759 1.14%
Online unproctored	46/1645 2.80%	52/1611 3.23%	30/1535 1.95%

Note. The data for the exams over two hours for the software proctored exams and for the online unproctored exams were eliminated from the respective data sets as invalid, while the data from the testing center were kept in the data set since the data were still valid.

Again, it is important to note that the tests in the testing center did not have time limits other than the limits on the hours the testing center was open. Therefore, data for students who took more than two hours in the testing center were still valid and were kept. This was somewhat an unequal factor in the time, but the number of tests over two hours in the testing center was small and should not have a big effect on the analysis.

The hypotheses related to the research questions were set forth as listed. Each hypothesis is listed with a number and a subscript to indicate the matching question. The null hypotheses are indicated with *0* for the subscripts and the alternative hypotheses with *a* for the subscripts.

$H1_0$: Mean exam scores for unproctored online students are not significantly higher than mean exam scores for software-proctored on-campus students.

$H1_a$: Mean exam scores for unproctored online students are significantly higher than mean exam scores for software-proctored on-campus students.

$H2_0$: Mean exam scores for unproctored online students are not significantly higher than mean exam scores for physically proctored on-campus students.

$H2_a$: Mean exam scores for unproctored online students are significantly higher than mean exam scores for physically proctored on-campus students.

$H3_0$: The mean time taken on exams by unproctored online students is not significantly higher than the mean time taken on exams by software-proctored on-campus students.

$H3_a$: The mean time taken on exams by unproctored online students is significantly higher than the mean time taken on exams by software-proctored on-campus students.

$H4_0$: The mean time taken on exams by unproctored online students is not significantly higher than the mean time taken on exams by physically proctored on-campus students.

$H4_a$: The mean time taken on exams by unproctored online students is significantly higher than the mean time taken on exams by physically proctored on-campus students.

The p values were to be compared to the α level of .05, which is a common standard (Sullivan, 2013). If the p value was less than or equal to .05, then the corresponding null hypothesis was rejected. Rejecting a null hypothesis would mean there was enough evidence to assume the mean of the exam scores or the mean of the time spent on the exams, depending on the test, was significantly different for the specific comparison.

Data Analysis

For the analysis of this study, the data was first cleaned in Excel, removing the scores and times that were invalid. This was mostly the removal of those online score-time pairs where the time was over two hours. Once the data was cleaned, a copy was kept in Excel, and a copy was made into a comma-separated values (CSV) file for use in Statistical Package for the Social Sciences (SPSS).

The analysis for this study was done both in Excel and SPSS. Excel was used mostly to look at the data from a graphical standpoint, including scatterplots and box and whisker plots. However, to answer the full question of statistical significance, as well as to address the research questions, the data were compared in SPSS. The method used was independent *t* tests, comparing each of the proctoring assessment methods (software and testing-center) individually against the unproctored online assessments.

The comparison was made with the unproctored online exams against each of the proctored forms of assessment using the independent *t* test for each of the three exams. This was done for both the scores and time taken on the exams. This created two *t* tests for the scores and two *t* test for the time on each unit exam. The results were twelve tests, six comparisons for exam scores, and six comparisons for the time taken on the exams.

Because there was no guarantee the data had equal variance, the results from SPSS for equal variance not assumed was used. The *p* values were compared to the .05 level to determine whether or not the difference was significant. If the *p* value was less than .05, the difference in the mean values for that test would be considered significant. If the *p* value was not less than .05, the mean values for that test would not be considered significant.

Each null and alternative hypothesis corresponds to three *t* tests on the related exams. To reject the null hypothesis, the *p* value needed to be less than .05 on all three tests. If any one of the three tests related to the specific null and alternative hypotheses was not less than .05, then that null hypothesis could not be rejected.

Results

Table 4 shows the means and standard deviations for each of the three exam scores under each of the three testing methods taken from SPSS.

Table 4

Means and Standard Deviations for Exam Scores

Testing type	Exam 1	Exam 2	Exam 3
Testing center	$M = 79.59$ $\sigma = 13.44$	$M = 77.06$ $\sigma = 15.84$	$M = 81.50$ $\sigma = 13.72$
Software proctored	$M = 80.32$ $\sigma = 13.35$	$M = 79.25$ $\sigma = 14.94$	$M = 83.69$ $\sigma = 12.38$
Online unproctored	$M = 80.40$ $\sigma = 14.39$	$M = 77.13$ $\sigma = 15.28$	$M = 82.35$ $\sigma = 13.74$

Table 5 shows the means and standard deviations for the amount of time spent by students on the exams for each of the three exams and for each of the three testing methods. There are some issues that should be considered about the time on the tests. For both online methods, the time starts when the student opens the exam and stops when the student closes the exam. For students in the testing center, the time starts when the student checks in. The student then must find an open seat and start the exam. The process of finding a seat might take some time, depending on how busy and crowded the testing center is. The time does not end until the student checks the exam back in at the counter. If there was a long line to check out, extra invalid minutes could be added to the time value.

Table 5

Means and Standard Deviations for Time Spent on Exams

Testing type	Exam 1	Exam 2	Exam 3
Testing center	*M* = 47.34 σ = 17.88	*M* = 49.92 σ = 19.44	*M* = 52.82 σ = 19.59
Software proctored	*M* = 38.60 σ = 14.24	*M* = 38.92 σ = 15.12	*M* = 40.98 σ = 15.25
Online unproctored	*M* = 56.89 σ = 25.13	*M* = 60.34 σ = 27.45	*M* = 58.21 σ = 24.74

The mean amount of time was much greater for the online unproctored data set than for the other two. Online unproctored time spent on exams was especially higher when compared against software proctored exams. The standard deviations for the online unproctored students are much higher than in the other two testing methods. Figures 2, 3, and 4 show the box and whisker plots for comparison of the three testing methods for each of the three exams. In looking at the box and whisker plots for the exam scores, all three have a similar spread on the data.

Figure 2. Exam 1 scores box and whisker comparison.

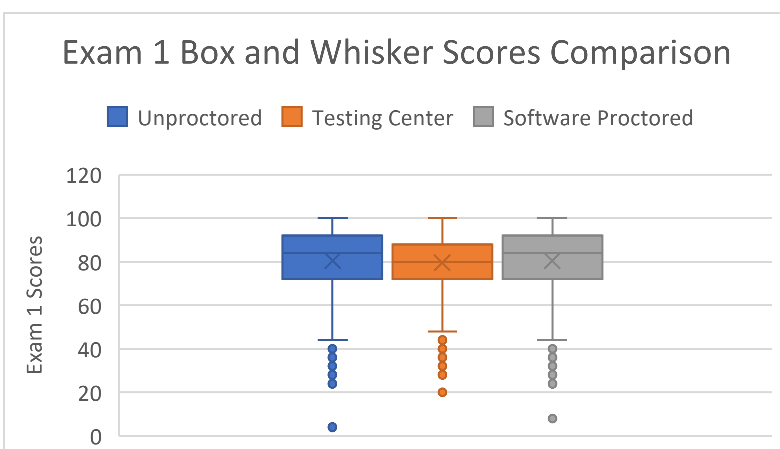


Figure 3. Exam 2 scores box and whisker comparison.

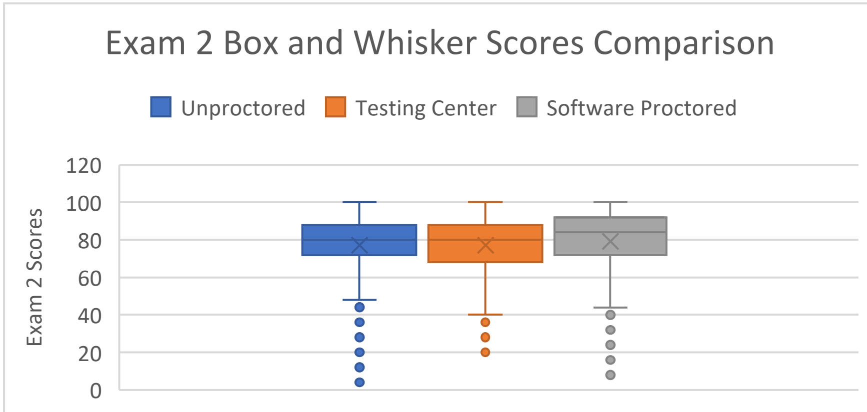
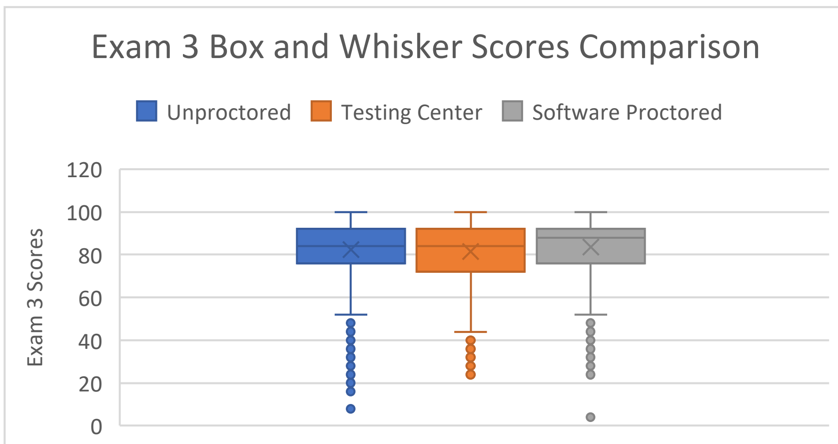


Figure 4. Exam 3 scores box and whisker comparison.



The box and whisker plots for the comparison of time in Figures 5, 6, and 7 show that the time taken on exams could have a much greater difference than the exam scores appeared to have.

Figure 5. Box and whisker plot comparison for times students spent on exam 1.

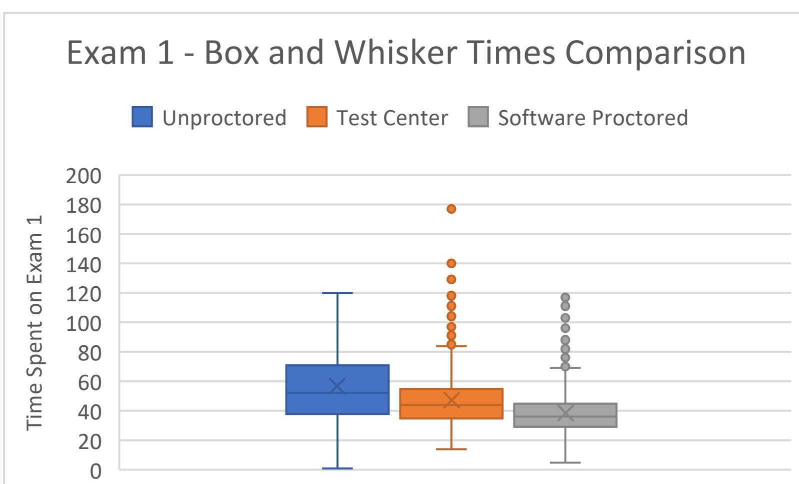


Figure 6. Box and whisker plot comparison for times students spent on exam 2.

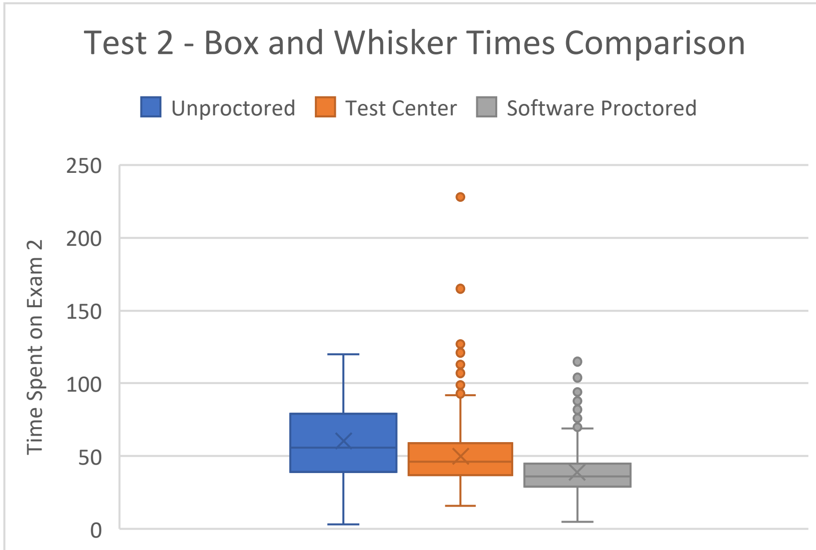
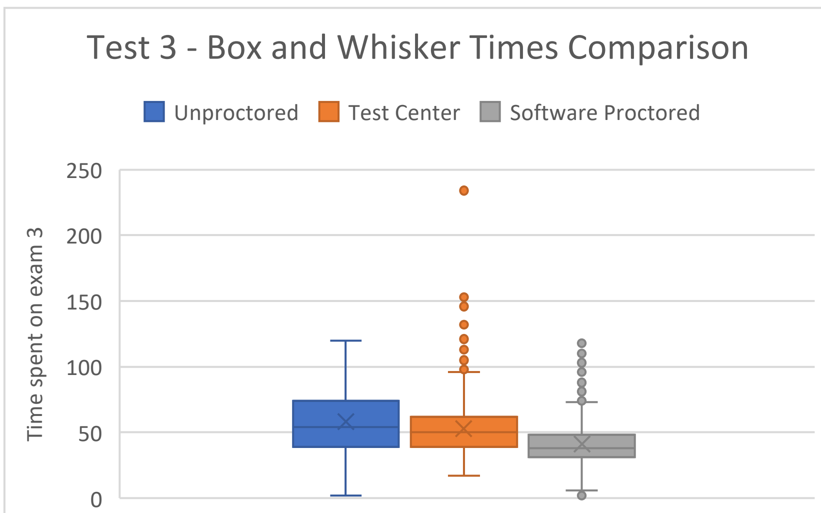


Figure 7. Box and whisker plot comparison for times students spent on exam 3.



The exam-scores box and whisker plots indicate the exam scores might not show a significant difference between the unproctored online classes and the two proctored forms of assessment. In contrast, the times taken on the exams appear to have more differences, with possibly significantly more time taken for unproctored online exams. A scatterplot of the three exams with a comparison of the three testing forms, using time as the x-axes and scores as the y-axes, gives an added dimension to what was happening in the examination process.

Figure 8 shows the scatter plot comparison of all three exams for all three forms of testing. (The four to six outliers above two hours for each exam in the testing center data were removed so the x-axis would be 130 for all three comparisons. The reason for doing this was twofold. First, all online assessment data over two hours was already removed because it was invalid. Second, there are few data points over two hours in the testing center, the data points over two hours are outliers, those few data points essentially do not change the data results, and with those data points added in the horizontal axis was stretched on the testing center graph and made it hard to see the comparison. Stretching all the unproctored and online proctored axes to match the testing center data horizontal axis compressed the data points in the graph into a smaller area, also making the comparison difficult to see. Keeping the axes the same width by removing the few outliers in the testing center data makes the comparison of what is happening much more obvious.) For both forms of proctored testing (the top and bottom sets) the data points appear thickest in the lower time amounts, with a few scattered points in the upper end. In contrast, for the unproctored exams (middle set), the examination points are quite uniform across much of the time, with a huge group right on the closing time of 120 minutes, meaning the students worked on the exam right up until the time limit when the exam closed.

Figure 8. Scatter plot time comparison on all exams for all three forms of testing. (The highest four to six outliers, being over 120 minutes, for the testing center were removed to keep the same x-axis width for easier comparison.)

All of this analysis shows there might not be much difference in exam scores between the three methods of testing. In contrast, with respect to time, the graphs show there could be a big difference in the comparisons between the times students spent on exams on both proctored exam groups compared to the times students spent on the unproctored exams.

The results of the data and the significance level of each proctored method compared to the unproctored testing are in Table 6 and come from SPSS. The *p* value significance levels shown in SPSS are for a two-tailed test, so the values needed to be cut in half for this research since the research uses one-sided tests. The values in Table 6 show the correct values for this research, which are half of what was shown in SPSS.

Table 6

P Value as Listed in Independent t *Test in SPSS* for Comparison of Exam Scores*

Testing type compared to online unproctored	Exam 1	Exam 2	Exam 3
Testing center	.063	.454	.060
Software proctored	.433	.00002[†]	.002

*The values are half of what is in SPSS since SPSS shows results for a two-tailed test. Even though the data for software proctoring on exam 2 and exam 3 show a *p* value less than the established .05, the means are reversed for the null hypothesis since software proctored exams means are higher than means for the unproctored exams.

[†]More accuracy was given here for exam 2 software proctoring than in the dissertation.

In considering these *p* values, none of the values are in a range for which the null hypotheses can be rejected (with more explanation on exams 2 and 3 for the software-proctored comparison given below). These *p* values correspond to two of the four research questions. The two related null hypotheses are:

$H1_0$: Mean exam scores for unproctored online students are not significantly higher than mean exam scores for software-proctored on-campus students.

$H2_0$: Mean exam scores for unproctored online students are not significantly higher than mean exam scores for physically proctored on-campus students.

For null hypothesis $H1_0$, the *p* value for exam 1 on the software proctored comparison was above the .05 level. Though the exam 2 and exam 3 comparisons show a *p* value less than .05, according to Table 5 on the software-proctored exams, the means of the software proctored exams were higher than the means of the online unproctored exams, instead of the other way around. Together, the two results mean none of the three exams give credible evidence for rejecting the null, so the null hypothesis $H1_0$ cannot be rejected. For the null hypothesis $H2_0$, all *p* values are above the established .05 level, so the null hypothesis $H2_0$ cannot be rejected. The indication is no claim can be made for the online unproctored mean exam scores being higher than the mean scores for either the testing center exams or the software proctored exams. In fact, on exams 2 and 3, the software proctored students had significantly higher exam scores.

For the research questions regarding time, each of the two proctoring exam methods was compared individually against the unproctored exam. This comparison was made for each of the three exams separately using SPSS. The *p* values for the time comparisons for each proctored method against the unproctored testing can be seen in Table 7, with significant digits shown.

Table 7

P Value as Listed in One-Tailed t *Test in SPSS* With Showing Significant Digits*

Testing type compared to online unproctored	Exam 1	Exam 2	Exam 3
Testing center	$1.17245 * 10^{-31}$	$4.78805 * 10^{-31}$	$2.79215 * 10^{-10}$
Software proctored	$8.6735 * 10^{-133}$	$1.32605 * 10^{-148}$	$4.8729 * 10^{-112}$

*These results are half of what shows in SPSS since this is a one-tailed test, and SPSS shows results for a two-tailed test.

These values have a far stronger significance than a *p* value at the .05 level. There are two of the four research questions to which the *p* values relate. The related null hypotheses which can be rejected are listed below, followed by the two corresponding alternative hypotheses to replace the null hypotheses.

$H3_0$: The mean time taken on exams by unproctored online students is not significantly higher than the mean time taken on exams by software-proctored on-campus students.

$H4_0$: The mean time taken on exams by unproctored online students is not significantly higher than the mean time taken on exams by physically proctored on-campus students.

$H3_a$: The mean time taken on exams by unproctored online students is significantly higher than the mean time taken on exams by software-proctored on-campus students.

$H4_a$: The mean time taken on exams by unproctored online students is significantly higher than the mean time taken on exams by physically proctored on-campus students.

Part of the reason that this study used both a physical proctoring method and a software proctoring method for comparison was to make sure the proctoring method was not the issue causing the difference in exam scores or the amount of time spent by students on exams. In both cases, the proctoring methods failed to reject the null hypotheses for the exams scores, and in both cases, the proctoring methods rejected the null hypotheses for the times spent by students on the exams. These results give a strong indication there was no significant difference in exams scores when students were not proctored compared to proctored, but there was significantly more time spent on exams by unproctored students compared to proctored students no matter what type of proctoring was used. Table 8 shows the percentage more time that was spent on the unproctored exam compared to each proctored method for each of the three exams.

Table 8

Percentage More Time Spent on Exams for Unproctored Compared to Each Proctored Method

Testing type	Exam 1	Exam 2	Exam 3
Testing-center-proctored	20%	21%	10%
Software-proctored	47%	55%	42%

One final question to consider was: How do the two proctoring methods compare to each other as far as both time and exam scores are concerned? Though a comparison of the two proctored testing forms was not the specific question of this research, the question is related and does deserve a brief answer. Table 9 shows the *p* value for a comparison of the two proctoring methods for scores and times.

Table 9

P *Values in SPSS* for Comparison Between Software Proctoring and Testing Center Proctoring*

Item analyzed	Exam 1	Exam 2	Exam 3
Scores	.073	$8.85*10^{-5}$	$9.5*10^{-6}$
Time	$1.523 * 10^{-43}$	$1.642 *10^{-56}$	$5.743 *10^{-61}$

*These results are half of what shows in SPSS since this is a one-tailed test and SPSS shows results for a two-tailed test. In all cases, testing center times were higher, while software proctoring exam scores were higher.

When comparing the two methods against each other in independent *t* tests, the *p* values for the time taken by students on exams come up less than .05 for all three exams, indicating the amount of time taken by students for the two proctoring methods was significantly different for all three exams. Looking at Table 5, software proctoring has a mean less than the mean for testing-center proctoring. The time difference between the two proctoring methods was 8.4 minutes on the first exam, 10.48 minutes on the second exam, and 11.27 minutes on the third exam. The *p* values indicated the time spent by students on software-proctored exams was significantly less than time students spent by students on testing-center-proctored exams.

As mentioned previously, the reason for these results might be the nature of the start and end times for the testing methods. For a software-proctored exam, the instant the exam time starts, the student is presented with the first exam question, and the instant the student finishes the last question, the time ends. The testing-center-proctored student's time starts at check-in. The student then must find a desk, which might take some time if the testing center is full. The time for the testing-center-proctored student does not end until the student checks out, and the checkout process could take some time if there is a line at the checkout desk. The differences in the times between the two testing methods increased throughout the semester. This increase in differences might be due to an increase in wait times for students at the testing center because of more students taking tests later in the semester.

In the independent *t*-test comparisons of exam scores, exam 1 shows a *p* value of .073, which was above the .05 threshold. The indication from the *p* value for exam 1 was the means are not significantly different between the two testing methods. The opposite was true for the scores of exams 2 and 3, which show *p* values less than .05, indicating there was a significant difference in the exam scores between the two methods for those exams. An examination of Table 4 shows, in every case, the exam scores are higher for software-proctored exams than for testing-center-proctored exams. The analysis of the reasons for this difference is beyond the scope of this study and is left for further research. If software proctoring does give students a significant advantage, perhaps because of the solitary space the student chooses to take the exam or for some other reason, maybe the reason the first exam shows no significant difference is because the testing process was new for most students, and the students had to become familiar with the software-proctoring testing method. The possibility exists, as students and educators become more familiar with software proctoring that other differences might be found between the two proctoring methods.

Discussion

Most similar research showed that both exam scores and the amount of time spent on the exams were greater for the online unproctored students (Alessio et al., 2017; Daffin & Jones, 2018). In this study, the exam scores were not found to be significantly different. In contrast to the exam scores in this study, the amount of time spent on the exams in this study was found to be significantly more for online unproctored students than for either testing-center-proctored students or software-proctored students.

One reason the exam scores might not have been different, even though the times were different, could be the nature of the testing-center proctoring and proctoring software, along with the exams and preparation provided. In this study, the students knew far ahead of taking the exam if the exam would be proctored. Students taking tests in the testing center knew there would be proctoring. For the online-proctored students, the link into the exam also informed the student if the exam was proctored, something the proctoring software inserted into the hyperlink to the exam. The university also required the software-proctored students be given a practice exam using the proctoring software, which alerted the students to the fact the exams were going to be proctored. In contrast, in one of the software proctoring studies, the students did not know the exam was being proctored until the students started the exam (Alessio et al., 2017).

Students entering an exam knowing the exam is going to be proctored would likely study differently from students who knew there was a chance to look things up while taking the exam. Proctored students would likely prepare more diligently, knowing looking items up while taking the exam would be flagged as cheating. The increased amount of studying could cause the students being proctored to have exam scores equal to the exam scores of the students who were unproctored, even if the unproctored students were able to look up answers and the proctored students could not.

In correlation with these ideas, the difference of this study compared to the other software proctored studies indicated the lack of significant difference in exam scores could be more than just the ability to look things up while taking the exam. The difference might be attributed to the difficulty level or the type of exam, which might be a determining factor in how much benefit was gained by the unproctored students' ability to look up answers. In the second study using software proctoring, the students knew which exams were being software-proctored ahead of taking the exams, just as in this study. Yet, the student still had much lower scores when proctored compared to the same student's scores when unproctored (Daffin & Jones, 2018). In the Daffin and Jones study, when the students took the proctored exam, even going into the exam knowing the exam was proctored, the means ranged from 57.74% to 74.91%, while the unproctored students ranged from 73.43% to 82.90%.

In contrast, the means in this study ranged from 79.25% to 83.69% for the software-proctored students, 79.66% to 81.58% for the testing-center-proctored students, and 77.13% to 82.35% for the unproctored students. The higher averages in this study compared to the study by Daffin and Jones likely indicate that the exams in this study were much easier for the students. More challenging exams might show a greater benefit in the exam scores from being unproctored and students being able to look up answers.

Varble's graph in Figure 1 gives credence to the idea of this study that suggests lack of proctoring makes cheating easier depends on the type of exam (Varble, 2014). According to Varble, when the items on the exam require remembering, there appears to be a bigger increase in the exam score when a student is unproctored, as compared to tests that fall into the categories of understanding, analyzing, and applying. Understanding, analyzing, and applying type exams are closer to skills testing, while remembering is more of a memorization assessment.

Both of the two assumptions, students knowing ahead of time that proctoring would be in place and the difficulty level and type of the test questions affecting the effectiveness of proctoring, are not fully known. These two assumptions are also key ideas for future research. Neither assumption was isolated in this research, and to verify the truthfulness of either one, each would need to be tested individually with confounding variables eliminated as much as possible.

Though Daffin and Jones (2018) did not describe in detail the nature of the exams which were given, the fact that the exams were psychology exams could mean there were many terms and definitions to know. If that were the case, the exams would be more of a remember-type assessment. The exams in this study, in contrast, had a mixture of terms, concepts, and calculations.

The results of the study by Hollister and Berenson (2009) bore out the idea the material being tested could be a big factor in whether or not proctoring makes a difference. In Hollister and Berenson's study, the students did computer skills testing on Word, Excel, and Access. The course dealt with learning how to use these software tools. When comparing the exam scores for students who were proctored in a physical environment compared to students who were not proctored, there was no significant difference. The indication of Hollister and Berenson's study compared to this study was that skill concepts are not something that would show a large difference in exam scores based on whether the students are proctored or not. Some reasons exam scores could differ based on proctoring are the challenge in remembering items.

As an example, suppose an exam is a sample from 100 random facts that must be remembered. Being in a situation where the person taking the exam could look up answers would probably make an immense difference in how well the student performed on the exam, because they could just look up the definitions asked on the exam. On the other hand, if the student had to learn how to do a certain type of math calculation, like solving an equation, take a derivative, or do an integral, a reasonable assumption is if the person did not have the skills necessary to perform the calculation, attempting to look up answers would be only minimally effective unless the exact problem could be found. A similar problem might give an idea or a reminder on how to solve the problem, but without the skill set necessary for solving the skill-based type of problem, the likelihood of solving the problem would be small.

There is literature that supports the importance of assessment design in reducing cheating (Egan, 2018), and this study backs that conclusion. Although assessment design is not the emphasis of this research, the results indicate some possible connections. Assessment design regarding proctoring is another area of importance for further research.

If the types of questions matter as suggested, it is important to know what the questions in this study looked like. For the exams in this study, some of the questions were remember type, but most of the questions were calculation or understanding based, and would fall more in the understand, analyze, and apply categories. Some examples of questions, taken from a practice test, which were similar to questions on the real exams, are given here.

Sample Practice Questions for the University GE Math Test

Question 1: Mary wants to show the populations of the five most populous states. Which of the following charts or graphs would be most appropriate to represent this information?

(a) Scatter Plot (b) Bar Chart (c) Time Series (d) Pie Chart
(e) A graph would not be appropriate for this data

Question 2: Identify the type of system of equations and solve the system of equations if possible.

$5x - 7y = 7$
$10x - 14y = 14$

Select the type of system and most correct meaning below.
(a) An independent system, the lines will cross in only one place so there is only one solution.
(b) An inconsistent system, the parallel lines will never cross so there is no solution.
(c) A dependent system, the parallel lines will never cross so there is no solution.
(d) An inconsistent system, there is really only one line so any of the infinite number of points on that line is a solution.
(e) A dependent system, there is really only one line so any of the infinite number of points on that line is a solution.

Question 3: In Tasheena's Anthropology class Quizzes are worth 15% of the final grade, Exams are worth 55%, Projects are worth 25%, and attendance is worth 5%. At midsemester Tasheena scored 123 out of 150 points on quizzes, 77, 85, and 91 on the first three exams. She got extra credit on her project with a score of 27 out of 25 possible points, and she had perfect attendance to class. Compute Tasheena's grade percentage in the class so far.

(a) 90.68 (b) 93.68 (c) 62.75 (d) 85.68 (e) 77.18

Question 6: The Grand Teton, a mountain on the Idaho/Wyoming border, is 13,776 feet above sea level. Convert the Grand Teton's elevation into miles. (Hint: 1 mile = 5280 feet.)

(a) 0.22 miles (b) 72,737,280 miles (c) 1.56 miles (d) 2.61 miles (e) 14.31 miles

Question 7: The final price of the textbook for a math course is 6% more than the listed price when tax and shipping costs are included. If the listed price is $149, what is the final price of the textbook? (Round to the nearest dollar.)

(a) $9 (b) $141 (c) $158 (d) $2483 (e) $167

As can be seen from these questions, on many of the questions, the student will need to understand and apply the material learned. It would do the student little good to attempt to look up these questions during the test. However, as mentioned previously, the idea that the type of question might affect scores and time spent on exams is an idea that needs further research.

Another question this study considered is whether the proctoring method made a difference in comparison to unproctored exams. As far as the research questions are concerned, the unproctored exams showed no significantly higher scores over either software-proctored exams or physically proctored exams. However, unproctored exams did show significantly more time on exams compared to both software-proctored exams and physically proctored exams.

In this study, the GE math course at the university is meant to be a freshmen level course and consists mostly of freshmen who have recently graduated from high school. There was a big difference between the on-campus sections and the online sections regarding maturity, as seen in Table 1. Even though both had a large proportion of students who were freshmen who had recently graduated from high school, Table 1 indicates the online sections had more students who were older and either coming back to school or coming to college for the first time as more mature students. There were fewer freshmen in the online sections. This maturity difference might have had a significant effect on the study (Ladyshewsky, 2015; Richardson & North, 2013), but would have to be considered separately in future research by looking in more depth at the student demographics.

With the issue of maturity set aside, the study by Richardson and North (2013) does possibly give some correlation with this study. The first correlation is that whether an exam is proctored or unproctored does make a difference, and the second correlation is that proctoring is not the only factor playing into the difference in exam scores. One of the reasons 3 out of 22 exams in their study did not show a difference in exam scores between the proctored and unproctored settings could be maturity, but another reason could be the form or content of the exams.

One concern about the validity of this study has to do with the fact the proctored students, both software-proctored and testing-center-proctored, were on-campus students, and the unproctored students were all online students. There was a question as to whether the delivery method could have made the difference. The best-case scenario would have been to have had all of the students be from the same group. Not having all of the students from the same group might be an issue that could be addressed in future research. However, some past research has attempted to look at this issue and indicates that any difference in the amount of cheating was more likely a matter of whether or not the assessment was proctored than the delivery method (Varble, 2014). However, more research is still needed in this area.

If the analysis from this research holds true, and the type of exam can be a determining factor in the need for proctoring, educators might wonder if there is any need to proctor skill-based exams. Considering the concern about impersonation raised in the literature, there is still a question as to whether the person taking the exam is the person who is registered for the class. The depth of monitoring might be reduced in the examination process for a skills-based test, but at least a validation of identity should exist in every major assessment.

Limitations and Future Studies

This study has a few limitations. First, the groups were not identical, raising concerns about the demographics of the groups. Two big differences were the online students had a lower percentage of freshmen, and the online students were an average of fifteen or more years older. These differences could cause a significant inequality in the data results. Further research is needed to determine the effects of the demographics on assessment outcomes. Also, the sampling method for the groups was not completely random, though true randomness is seldom possible in a study of this type (Stack, 2015).

In contrast to other similar studies, this study came out with similar exam scores for both proctored and unproctored exams, even though the mean amount of time taken on the unproctored exams was significantly greater. In an attempt to understand why this study came out so much different, a few assumptions were made. These assumptions were:

1. The amount of preparation material provided to the students could have made a difference.
2. The type of questions on the exams, as suggested by Varble, could have made a difference (Varble, 2014).
3. Knowing in advance whether the test was proctored could have made a difference.
4. The difference in demographics of the students might have made a difference since both of the proctored groups were on-campus students, and the unproctored group was a mixture of purely online and on-campus students. Part of the demographic difference was age and class (freshmen, sophomore, junior, senior), which was different for online classes compared to on-campus classes. The way the registering works could also have affected the demographics of the online course.

Each of these assumptions has little or no research outside of this study to verify it and needs further research. One other limitation of this study is the nature of the university where the study was performed. BYUI is a religious school, and the results may not apply to other universities.

Conclusion

In summary, the analysis of this research provides some insights into the effects of proctoring on student assessments in the GE math course at the university. The research questions related to two things: (a) an expectation the mean of the scores on unproctored exams would be higher for unproctored students compared to proctored students, and (b) the expectation the average amount of time students spent on exams would be higher on unproctored exams than on proctored exams. The hypothesis was that unproctored students would be able to cheat more easily than proctored students, and the cheating would show up in the higher exam scores and the increased amount of time taken by the students on the exams.

This research clearly showed that unproctored students spent a significant amount more time on exams than students did who were proctored. This increase in time did not translate into a significant increase in exam scores, and the reason is probably threefold. The first reason is the students who were proctored knew in advance of taking the assessments the proctoring would be in place. This foreknowledge was different from one study where both the mean on exam scores and the amount of time students spent on exams was higher for unproctored exams (Alessio et al., 2017). Knowing the assessments would be proctored, the assumption is the students spent more time in preparation for the assessment, making the difference in the exam scores insignificant.

The second reason there was likely no significant difference in exam scores between those who were proctored and those who were not, even though the time values were significantly different, is somewhat connected to the first. The reason is because the nature of the exams meant looking answers up could slightly help a person who had not prepared well, but the concepts could still be readily mastered by someone who took the time to prepare.

The third reason was the types of questions on the exam. The research by Varble indicates that remember-type questions are easier for students to cheat on than are questions of types understand, analyze, and apply (Varble, 2014). The exams in this study had a mixture, but a large proportion were questions of the understand, analyze, and apply types, possibly making looking up answers during the exam less beneficial for the students.

Although all three of these assumptions need to be verified with further research, there is some research that suggests the assumptions might be valid. In one study where students knew ahead of time the assessment would be proctored, the students still scored lower on the exam score, even though the amount of time taken on the exams was higher for unproctored students (Daffin & Jones, 2018). The difference between the Daffin and Jones (2018) study and this study might be the type of material on the exam, though that, too, needs more verification.

Studies indicated that something like a skills test might not show a significant difference between exam scores for proctored and unproctored exams (Hollister & Berenson, 2009). The nature of the exams and the material on the exams might play a big part in how much being able to look up answers could affect the grade (Egan, 2018). Exams requiring memorization components might be affected the most by proctoring versus not proctoring (Varble, 2014). The differences in the types of material on the exams in this study compared to the exams in the Daffin and Jones study could be the reason the exam scores were not significantly different between proctored and unproctored exams in this study, yet were significantly different in the Daffin and Jones study.

This information, taken together with what previous studies have shown, indicated the need for proctoring likely has some key components. The first component is the nature of the assessment. If the assessment is skills or analysis based, the need for proctoring might be less than if the assessment is memory-based or has a large memory component. Another consideration is that if the students are more mature, there might be less need to proctor (Ladyshewsky, 2015; R. Richardson & North, 2013). Finally, if the assessment has information which is not easy to look up, the assessment may not have as much need for proctoring as an assessment which has material which can easily be found online. Outside of all other factors for determining whether an assessment should be proctored or not, one of the greatest concerns should be the possibility of impersonation (Daffin & Jones, 2018). In deciding whether an assessment needs to be proctored, the concern about impersonation should be considered. If the assessment has a high possibility of having impersonation, the assessment should be proctored even if the other indicators for the need for proctoring are low.

Acknowledgments

The author thanks his committee Dr. Dennis Debay and Dr. Mary Strong, and American College of Education faculty Dr. Elizabeth Johnson, Dr. Imani Akin, Dr. Crystal Neumann, and Dr. Ted Creighton. Also, the author thanks colleagues at BYUI who helped with approval and data collection and review of the analysis, Jennifer Gohr, Elaine Wagner, Danae Romrell, Jackie Nygaard, Jason Rose, Ryan Cromar, Craig Johnson, Eric Karl, Curtis Henri, Seth Demming, Austin Lee, Dane Bohman, Jim Croasman, Melanie Kenelly, and Sid Palmer.

References

110th U.S. Congress. (2008). Higher Education Opportunity Act. *Public Law*. http://www.gpo.gov/fdsys/pkg/PLAW-110publ315/pdf/PLAW-110publ315.pdf

Alessio, H. M., Malay, N., Maurer, K., Bailer, A. J., & Rubin, B. (2017). Examining the effect of proctoring on online test scores. *Online Learning*, *21*(1), 1–17. https://doi.org/10.24059/olj.v21i1.885

Bemmel, M. B. (2014). *Cheating in online classes: A preliminary investigation* [Doctorl dissertion]. Nova Southeastern University. https://nsuworks.nova.edu/fse_etd/37/

Berkey, D., & Halfond, J. (2015). Cheating, student authentication and proctoring in online programs. *New England Journal of Higher Education*, *July 20*, 1–4. http://www.nebhe.org/thejournal/cheating-student-authentication-and-proctoring-in-online-programs/

Brown, V. (2018). Evaluating technology to prevent academic integrity violations in online environments. *Journal of Distance Learning Administration*, *21*(1), 1–12. https://www.westga.edu/~distance/ojdla/spring211/brown211.html

Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., & Thies, W. (2015). Deterring cheating in online environments. *ACM Transactions on Computer-Human Interaction*, *22*(6), 1–23. https://doi.org/10.1145/2810239

Daffin, L. W., & Jones, A. A. (2018). Comparing student performance on proctored and non-proctored exams in online psychology courses. *Online Learning*, *22*(1), 131–146. https://doi.org/10.24059/olj.v22i1.1079

Egan, A. (2018). *Improving academic integrity through assessment design.* Dublin City University, National Institute for Digital Learning.

Fendler, R. J., Ruff, C., & Shrikhande, M. (2018). No significant difference—unless you are a jumper. *Online Learning*, *22*(1), 39–61. https://doi.org/10.24059/olj.v22i1.887

Hollister, K. K., & Berenson, M. L. (2009). Proctored versus unproctored online exams: Studying the impact of exam environment on student performance. *Decision Sciences Journal of Innovative Education*, *7*(1), 271–294. https://doi.org/10.1111/j.1540-4609.2008.00220.x

Howard, D. (2019). *Online testing integrity in a general education math course: A correlational study*. American College of Education. https://www.proquest.com/openview/b0235bd6a6ffa3d3864cdd00a8848d3c

Ladyshewsky, R. K. (2015). Post-graduate student performance in "supervised in-class" vs. "unsupervised online" multiple choice tests: Implications for cheating and test security. *Assessment and Evaluation in Higher Education*, *40*(7), 883–897. https://doi.org/10.1080/02602938.2014.956683

Madara, D. S., & Namango, S. S. (2016). Faculty perceptions on cheating in exams in undergraduate engineering. *Journal of Education and Practice*, *7*(30), 70–86. http://www.iiste.org/Journals/index.php/JEP/article/viewFile/33621/34565

NCES. (2016). *Digest of education statistics: 2016*. https://nces.ed.gov/programs/digest/d16/

Polat, M. (2017). Why do students cheat in examinations in Turkey? A meta-synthesis study. *International E-Journal*, *7*(1), 203–222. https://doi.org/10.22521/jesr.2017.71.7

Ravasco, G. G. (2012). Technology-aided cheating in open and distance e-learning. *Asian Journal of Distance Education*, *10*(2), 71–77. https://www.academia.edu/2092903/Technology-Aided_Cheating_in_Open_and_Distance_e-Learning

Richardson, R., & North, M. (2013). Strengthening the trust in online courses: A common sense approach. *Journal of Computing Sciences in Colleges*, *28*(5), 266–272. https://dl.acm.org/citation.cfm?id=2458624

Sleight, R. (2020, August 2020). Personal communication to author.

Stack, S. (2015). Learning outcomes in an online vs traditional course. *International Journal for the Scholarship of Teaching and Learning*, *9*(1), 1–20. https://doi.org/10.20429/ijsotl.2015.090105

Sullivan, M. (2013). *Statistics: Informed decisions using data* (3rd edition). Pearson.

Varble, D. (2014). Reducing cheating opportunities in online tests. *Atlantic Marketing Journal*, *3*(2), 131–150. https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1110&context=amj

Watson, G., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Learning Administration*, *13*(1), 798–803. http://www.westga.edu/~distance/ojdla/spring131/watson131.html