IMPACTS OF COVID-19 ON EXAMINATION DURING THE SUDDEN SWITCH TO ONLINE IN THE SPRING AND FALL OF 2020

Jeffrey K. Landgren
University of North Georgia
Dahlonega, GA 30597
jeffrey.landgren@ung.edu

Gregg Velatini
University of North Georgia
Dahlonega, GA 30597
gregg.velatini@ung.edu

Introduction. Over the past decade online coursework has increased substantially. The motives for universities to provide online coursework include, but are not limited to, improving access and convenience. [1] With the increase in online teaching, it has become more important to evaluate the effectiveness of online learning. One tool for analyzing online efficacy begins with looking into examination records. However, with examination records it is imperative to sort the academic integrity of those grades. While this can be best analyzed using a large number of students under a controlled environment, the pandemic caused by COVID-19 has provided an opportunity to assess the integrity of online examination as a consequence of a large scale shift to online coursework under the specific circumstances laid out in this paper.

Generally speaking, online coursework has vast pros and cons that have been documented thus far. Some reports show that online coursework has a positive impact on completing a degree when relevant background characteristics are accounted for and yet others state that online coursework has a negative impact on both course grades and course persistence. [2–5] Furthermore, the perception of online coursework from the start can have a negative impact on a students' mindset entering a course conducted online. While switching to online provides easier access to education, the jury is out on its overall effectiveness when compared to in-person learning. Moreover, specific circumstances matter significantly. For the purposes of this study, the authors have chosen to use linear regression models, an ANOVA, t-tests, and chi-squared tests to quantify the ethics of the exams administered.

Academic dishonesty can occur by plagiarism, consulting with another student or set of materials outside of the list provided by the instructor, or when a different student takes an exam for a student enrolled in the course. Not surprisingly, since online coursework is on the rise, there is some discussion that so is academic dishonesty. [6, 7]

With respect to exams, many institutions maximize proctoring to combat academic dishonesty and often require identification at the time of the exam. Other schools employ honor codes in an attempt to mitigate academic dishonesty. For online exams, ProctorU, Remote Proctor Now, and other virtual proctoring platforms have been employed and have shown to be effective. [8]

During the year 2020, switching to online became imperative to continue education during the COVID-19 pandemic. For some professors little preparation time meant using exams typically used in the classroom for continued use online. Research of controlled environments typically show better performance by students of online exams when not proctored. [9, 10] In this paper, anomalies in grades are highlighted from the collection of exam scores found during 2020, usually being exams that were conducted online, and compared with those from previous years at the University of North Georgia that were given in-person using the programming language R.

Examination Formats. As previously stated, the focus of this project remains to observe anomalies that occurred in examination during the semesters that coincided with the pandemic. In particular, the authors of this paper felt enough data was observed in both the spring of 2020 and the fall of 2020 for an introductory statistical analysis.

Most faculty at the University of North Georgia in the spring of 2020 gave at least one exam that was proctored in-person. The latter exams in the spring semester were conducted online due to the closure of in-person instruction. For consistency, the authors of this article only analyzed exams from courses where exam one was proctored in-person and exam two, three, and the final were conducted online. As a result of the sudden switch to online many faculty provided exams that were of a different format. For example, most faculty in the mathematics department at the University of North Georgia provided multiple choice exams using Webassign or MyLabMath once the switch was made to online. Switching to online exams that are provided in a different format, makes sussing out anomalies too challenging. The analysis that follows is over exams that did not change format. The exams included in the analysis were given online and consisted of all free response questions, just as they would have been given during any normal academic year. Each online exam was conducted online through D2L, the university platform that allows students to access material, take quizzes and tests, and see grades.

On a day that an exam was administered the online exam was provided as follows. The students first downloaded a pdf of their exam, then wrote their solutions down on either the exam itself, after printing it, or on blank paper. Following this the students took a photo or photos of their solutions, converted those photos to a single pdf, and then uploaded their pdf back onto the D2L platform. Students were allowed to use a free app that converted their photos to pdf format for them. Additionally, the time given for these exams was fifteen minutes longer than normal to account for

the extra procedures. That is, students were granted fifteen minutes longer than they would have been given if the exam were conducted in person. The extra time was not only permitted so that the students could conduct the aforementioned process, but also to allow for the possibility that many students might be uploading to the platform at the same time. Days before the exam was conducted, students performed a graded practice run to get an idea of what the procedure would feel like. Having a test run allowed students to work through potential setbacks.

Exams during the fall of 2020 were conducted differently. In fact, the story almost flips entirely. Exam one, two, and three were conducted in person while the final exam was conducted online. Regardless, the exams in the fall of 2020 were given in the same format that was used for in-person exams, only free response questions, and the exams were given as described in the procedure mentioned with regard to the D2L platform.

Beginning of the Anomolies. The origin of the anomalies occurred first during the spring of 2020, but continued into the fall semester. Such examples throughout the year that sparked the discussion, which then resulted in the writing of this paper are laid out below.

	Exam 1	Exam 2	Exam 3	Final Exam
Student A	36	18	34	60
Student B	29	25	24	74
Student C	20	20	20	80

TABLE 1. Exam grades that occurred during 2020. Each grade is out of 100 total points. Exams seen in blue were conducted online.

TABLE 1 shows a sample of students who performed poorly on each exam and then substantially better in their online final exam.

Additionally, anecdotal information and box plots (FIGURE 1 seen on the following page) from a professor who conducted three Calculus III courses showed significant anomalies. The professor described the 2018 Calculus III course to contain the highest performing students they have had out of all of their years of teaching at the university. Notice exam 1, 2, and 3 averages sequentially get better for the 2018 course. However, the final exam average is significantly lower than all previous exams. Then, the opposite trend occurs in 2020. Exam averages consecutively become lower with an online final exam average higher than the in-person exam 2 and 3.

At first sight of the examination scores in the fall of 2020, an ANOVA was used to compare three different years of calculus I final exams, 2017, 2019, and 2020. The 2017 and 2019 finals were proctored in-person, while the 2020 online final was not

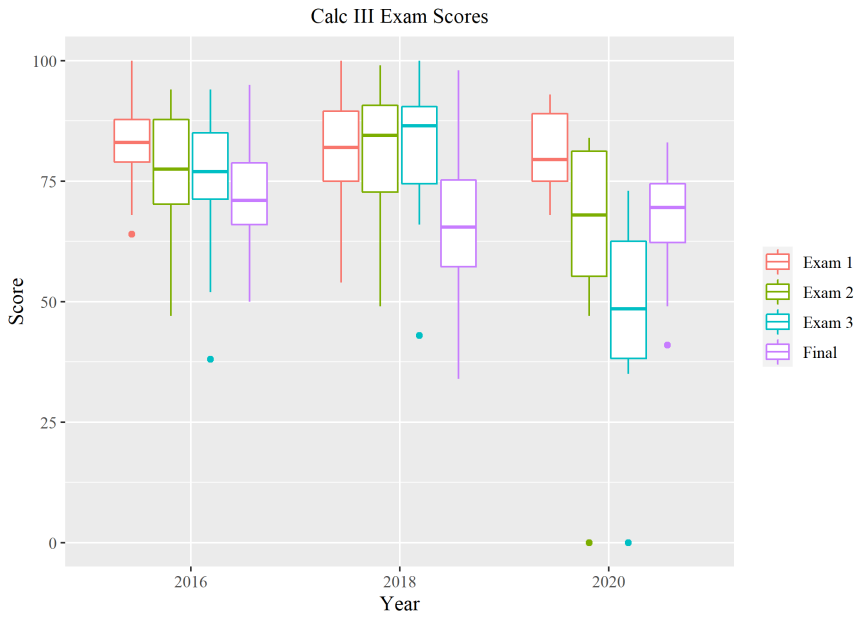

FIGURE 1. Box plots for three Calculus III courses.

proctored. The result was significant ($p = 3.12e{-}04$). Next a two-sample t-test was used to compare the 2017 and 2020 final exam scores. The average on the 2020 final exam was almost 20 points higher than the average of the 2017 scores. The result here was also significant ($p = 5.06e{-}05$). At this point, it became clear that further investigation was needed (Please see APPENDIX A for further information regarding the ANOVA and t-test).

MAIN BODY OF PAPER

Spring 2020. To begin, the courses that were used from the spring 2020 semester that were then compared to other courses from previous years were Calculus II, Pre-calculus, and Differential Equations. Recall that examinations were conducted during the spring time in the following manner:

Exam 1	Exam 2	Exam 3	Final Exam
In-person	Online	Online	Online

TABLE 2. Examination arrangement during the spring of 2020.

Results. With TABLE 2. in mind, a comparison was conducted of two quantities for the spring courses. The first quantity of interest is labeled $\mathcal{S}_1$ as seen in TABLE 3. $\mathcal{S}_1$ represents the number of students in a course who scored higher on two of the remaining three exams (Exam 2, Exam 3, and Final Exam) following the first exam.

Similarly, $\mathcal{S}_2$ is the number of students in a course who scored higher on all three of the remaining three exams (again Exam 2, Exam 3, and Final Exam).

Year	Class	Section	$\mathcal{S}_1$	$\mathcal{S}_2$	n
18	Calculus II	A	5	1	22
18	Calculus II	B	4	3	27
20	Calculus II	A	7	5	22
18	Differential Equations	A	21	7	29
20	Differential Equations	A	27	19	30
17	Precalulus	A	19	7	38
19	Precalulus	A	2	1	33
20	Precalulus	A	16	11	28
20	Precalulus	B	6	5	14

TABLE 3. $\mathcal{S}_1$ is the number of students that scored higher on 2 out of the remaining 3 exams. $\mathcal{S}_2$ is the number of students that scored higher on 3 out of the remaining 3 exams. n is the total number of students in the course. Highlighted in blue are the 2020 courses.

Seen in TABLE 3, the majority of the courses in 2020 have higher $\mathcal{S}_1$ values and $\mathcal{S}_2$ values. The exception is the 2017 Precalulus course. However, what must be taken into account is the class sizes. Notice that the 2017 Precalulus course includes 38 students whereas the 2020 Precalulus Section A course includes 28 students. Additionally, the $\mathcal{S}_2$ value, a more challenging feat than $\mathcal{S}_1$, is most often higher in 2020 than previous years.

Time	Total $\mathcal{S}_1$	Total $\mathcal{S}_0$
spring 2020	56	38
Prior to 2020	51	98

TABLE 4. $\mathcal{S}_1$ is the number of students that scored higher on 2 out of the remaining 3 exams. $\mathcal{S}_0$ is the number of students who did not score higher on 2 out of the remaining 3 exams.

Following the aforementioned observations a chi-squared test was run. The comparison was made from the data seen in TABLE 4. The chi-squared test was used to determine if being in the spring of 2020, versus being a student prior to 2020, was related to scoring higher on 2 of the remaining 3 exams. The independence test determined ($p = 1.81e{-}04$) that there is a relationship between test scores and whether they were taken in 2020 or before 2020. The year 2020 stands out with the Total $\mathcal{S}_1$ value being higher in 2020 than prior to 2020 and a Total $\mathcal{S}_0$ value smaller than prior to 2020.

Fall 2020. The courses that were used in computations that took place in the fall semester include Calculus I, Calculus III, and Precalculus. Let us recall how examination was conducted during the fall of 2020:

Exam 1	Exam 2	Exam 3	Final Exam
In-person	In-person	In-person	Online

TABLE 5. Examination arrangement during the spring of 2020.

Results. To begin the analysis for fall, a similar table to TABLE 3 was computed. Keeping TABLE 5 in mind, the final exam scores are compared with each in-person exam (Exam 1, 2, and 3):

Year	Class	Section	$\mathcal{F}_1$	$\mathcal{F}_2$	n
17	Calculus I	A	13	5	25
19	Calculus I	A	6	2	33
20	Calculus I	A	17	10	21
16	Calculus III	A	6	0	26
18	Calculus III	A	13	5	25
20	Calculus III	A	17	10	21
17	Precalulus	A	16	7	38
19	Precalulus	A	4	0	33
20	Precalulus	A	14	3	30
20	Precalulus	B	12	6	28

TABLE 6. $\mathcal{F}_1$ is the number of students who scored higher on their final exam than 2 out of 3 prior exams. $\mathcal{F}_2$ is the number of students that scored higher on their final exam than 3 out of the 3 prior exams taken. n is the total number of students in the course. Highlighted in blue are the 2020 courses.

This time around, the number of students who scored higher on their final exam than two out of three previous exams is recorded as $\mathcal{F}_1$. The number of students who scored higher on their final exam than all three previous exams is designated under the label $\mathcal{F}_2$. Similar trends to spring 2020 are observed in the fall of 2020. Most often the 2020 courses had higher $\mathcal{F}_1$ and $\mathcal{F}_2$ values despite their smaller class sizes.

Next a linear regression was conducted on three Calculus III courses. The regression compares the predictability that the average of exam 1, 2, and 3 have on the outcome of the final exam:

$$FinalExam = \beta_0 + \beta_1 \times (MidtermAverage) \tag{1}$$

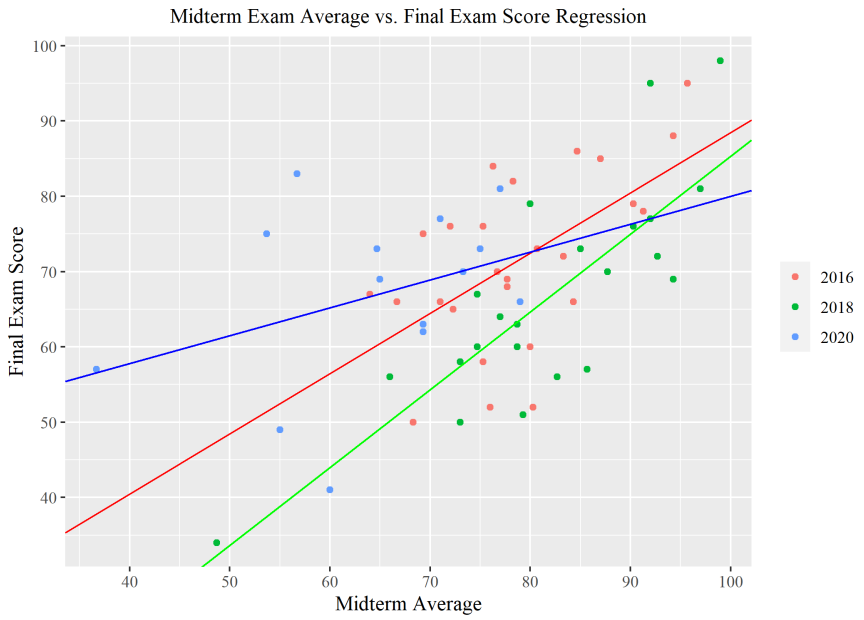

FIGURE 2. Linear Regression model on Calculus III courses where the x-axis represents the average of Exam 1, Exam 2, and Exam 3. The y-axis represents the Final Exam score.

Seen in FIGURE 2 the slopes of the 2018 and 2016 lines are almost parallel. Precisely, their slopes are 1.03 and 0.80, respectively. The 2020 slope is 0.37, quite a bit smaller. The smaller slope indicates that for 2020 the regression line will predict much higher final exam scores for students who have midterm averages of less than 75. In other words, during the year 2020, a student that scores a low midterm average is likely to score higher on their final exam. Looking at a particular data point that corresponds to 2016 or 2018 one can observe that the linear regression indicates going up one point along the midterm average predicts going up one point for a corresponding final exam grade. However, up one point on the midterm average means up by 3 points for a corresponding final exam grade if a student took the course in 2020.

Next multiple regression with a categorical predictor was conducted to determine the significance of the year 2020:

$$FinalExam = \beta_0 + \beta_1 \times (MidtermAverage) + \beta_2 \times (Year) \tag{2}$$

The model shows that the year mattered. In (2) let $Year$ equal 1 if the year is 2020 and let $Year$ equal 0 if the year is not 2020. The coefficient of the variable Year, β_2, is significant (p = 1.58e−02). The year 2020 is a significant predictor of final exam score. For the same midterm score, the year 2020 predicts a 9.39% higher final exam score than other years.

Conclusions

Student grades improved significantly if there was an online non-proctored exam throughout the semester. The average of each online non-proctored exam was higher than the average of each proctored in-person exam from a previous year. The online exams displayed trends that were not seen in the in-person proctored setting. These anomalies can be explained by the following list that include, but are not limited to, class sizes, the manner in which instruction was conducted, the difficulty of the material, more stress for students created by the pandemic, online instruction is more effective than in-person instruction, academic dishonesty, the content assessed on exams was different than in previous years, different textbooks, different instructors, small sample sizes, and the students had more time in online exams. It is the opinion of the authors that the following confounding variables do not show a strong causation of the anomalies: the manner in which the instruction was conducted, the difficulty of the material, the content assessed on exams was different than in previous years, different textbooks, and different instructors. All of these variables were equivalent or exactly the same from year to year. The authors recognize that this is an observational study, and therefore make no claims about the cause of the higher test scores. A randomized controlled experiment is needed to assess causation. However, academic dishonesty remains a strong possibility.

Appendix A. Additional Results

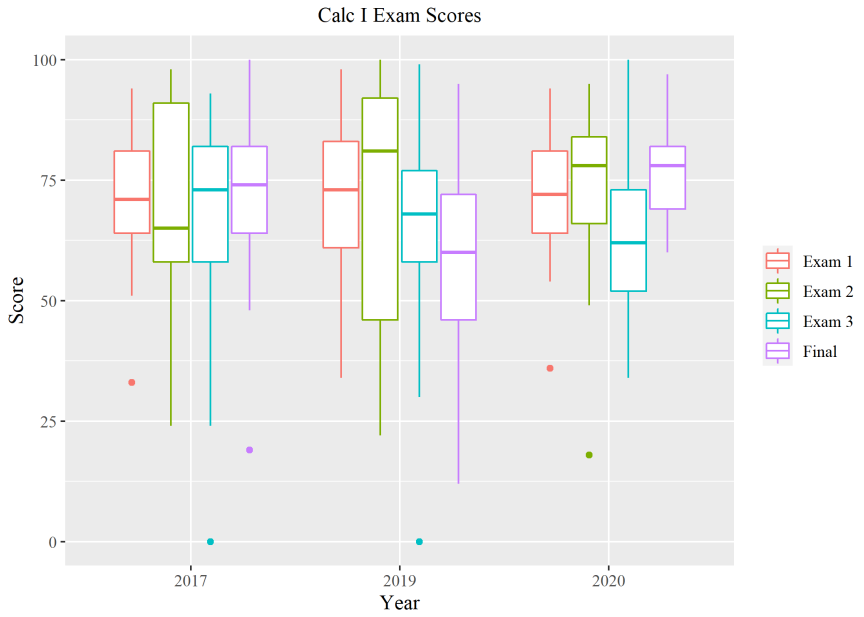

Figure 3. Box plots for three Calculus I courses.

An ANOVA was conducted comparing three Calculus I final exam scores after observing that the bottom quarters were up. The box plots associated with the Calculus I scores are seen in FIGURE 3. The null hypothesis for the ANOVA test states that all three groups of final exams are the same (come from the same population of students.). The probability that the groups come from the same population with the observed difference in final exams by random sampling from the single population is 0.03%. It is possible that a coincidence has occurred, but the coincidence only occurs 0.03% of the time. Thus, approximately 99.97% of the time the difference can be explained by the three groups coming from different populations. Furthermore, the test reveals that these three groups are not the same.

Also, two Sample t-tests were run to compare the 2020 final exam average to the 2019 and 2017 final exam average. The 2020 final exam average is about 20 points higher than the 2019 final exam average. Since the p-value is 5.06e−05, then the results show a significance at all levels ($p < 0.01e{-}01$).

```
##
##  Welch Two Sample t-test
##
## data:  scores_2019 and scores_2020
## t = -4.4371, df = 49.752, p-value = 5.056e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -27.96653 -10.53564
## sample estimates:
## mean of x mean of y
##  57.93939  77.19048
```

FIGURE 4. Sample t-test comparing Calculus I 2019 to 2020.

```
##
##  Welch Two Sample t-test
##
## data:  scores_2017 and scores_2020
## t = -1.5192, df = 41.97, p-value = 0.1362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.854987   1.954034
## sample estimates:
## mean of x mean of y
##  71.24000  77.19048
```

FIGURE 5. Sample t-test comparing Calculus I 2017 to 2020.

For the 2017 Calculus I course the final exam average was 5 points below that of the 2020 final. The Two Sample t-test results are not significant with a p-value of 1.36e−01 ($p > 0.10$). Having a small sample size should be noted here.

REFERENCES

[1] K. KIM, C. J. BONK, *The future of online teaching and learning in higher education: The survey says...*, Educause Quarterly 29(4), 2006, 22.

[2] P. Shea, T. Bidjerano, *Does online learning impede degree completion? A national study of community college students*, Computers and Education 75, February 2014, Pages 103-111.

[3] D. Xu, Shanna S. Jaggars, *The impact of online learning on students course outcomes: Evidence from a large community and technical college system*, Economics of Education Review 37, August 2013, Pages 46-57.

[4] J. Anstine, M. Skidmore, *A Small Sample Study of Traditional and Online Courses with Sample Selection Adjustment*, The Journal of Economic Education Vol. 36, No. 2, Spring 2005, Pages 107-127.

[5] R. Bernard, P. Abrami, Y. Lou, E. Borokhovski, A. Wade, L. Wozney, P. A. Wallet, M. Fiset, B. Huang, *How Does Distance Education Compare With Classroom Instruction? A Meta-Analysis of the Empirical Literature*, Review of Educational Research Vol. 74, No. 3, Fall 2004, Pages 379-439.

[6] L. R. Jones, *Academic Integrity & Academic Dishonesty: A Handbook About Cheating & Plagiarism*, Scholarship Repository at Florida Tech, 2011, Pages 1-28.

[7] D. L. R. Jones, *Acadeic Dishonesty: Are More Students Cheating?*, Business Communication Quarterly Vol. 74, No. 2, June 2011, Pages 141-150.

[8] T. H. Reisenwitz, *Examining the necessity of proctoring online exams*, Journal of Higher Education Theory and Practice Vol. 20, No. 1, 2020, Pages 118-124.

[9] L. W. Daffin, A. A. Jones, *Comparing Student Performance on Proctored and Non-Proctored Exams in Online Psychology Courses*, Online Learning Vol. 22, No. 1, 2018,Pages 131-145.

[10] D. Prince, R. Fulton, T. Garsombke, *Comparisons Of Proctored Versus Non-Proctored Testing Strategies In Graduate Distance Education Curriculum*, Journal of College Teaching & Learning Vol. 6, No. 7, November 2009,Pages 51-62.