Examining the Effect of Proctoring on Online Test Scores

Helaine M. Alessio, Nancy Malay, Karsten Maurer, A. John Bailer, and Beth Rubin
Miami University

Abstract

Online education continues to grow, bringing opportunities and challenges for students and instructors. One challenge is the perception that academic integrity associated with online tests is compromised due to undetected cheating that yields artificially higher grades. To address these concerns, proctoring software has been developed to address and prevent academic dishonesty. The purpose of this study was to compare online test results from proctored versus unproctored online tests. Test performance of 147 students enrolled in multiple sections of an online course were compared using linear mixed effects models with nearly half the students having no proctoring and the remainder required to use online proctoring software. Students scored, on average, 17 points lower [95% CI: 14, 20] and used significantly less time in online tests that used proctoring software versus unproctored tests. Significant grade disparity and different time usage occurred on different exams, both across and within sections of the same course where some students used test proctoring software and others did not. Implications and suggestions for incorporating strategic interventions to address integrity, addressing disparate test scores, and validating student knowledge in online classes are discussed.

Keywords: online education, academic integrity, online testing, proctoring software, online course grades

Introduction

A recent analysis of Integrated Postsecondary Education Data System (IPEDS) data stated that about 5.3 million students, representing more than 25% of total college enrollment, took at least one online class in 2013 (Allen & Seaman, 2015). The increased popularity of online classes presents benefits and challenges to students, faculty, and academic institutions. Geographic locations and time zones no longer present obstacles for students to enroll in a class since online classes can be delivered nearly anywhere in the world with an internet connection.

This provides students an opportunity to advance in their studies while working, traveling, and attending to family responsibilities. In this paper, data is presented for a popular online elective class with an initial enrollment of 180 students that compares proctored and unproctored tests within and across class sections. The effect of proctoring was large enough to suggest an impact on test time and scores.

Literature Review

The credibility of online classes faces criticism due to the distance between students and instructors that may contribute to breaches in integrity (Moten, Fitterer, Brazier, Leonard, & Brown, 2013). Researchers contend that online programs must address student integrity; the use of proctoring software is one way to do so, to try to assure that students are being fairly and effectively evaluated. Moten and colleagues explained that in online courses, students work in relative autonomy and anonymity and instructors may not be certain who is taking exams or how best to validate learning (2013). In addition, Berkey and Halfond (2015) have examined the sensitive subject of cheating in online courses, and found an alarming 84% of 141 students who responded to their survey agreed that student dishonesty in online test taking was a significant issue. Yet, less than half the students surveyed indicated they had ever used proctoring software in online tests.

In a study by King, Guyette, and Piotrowski (2009), 73% of 121 undergraduate students surveyed felt it was easier to cheat online compared to a traditional face-to-face classroom. When asked if they were likely to cheat, a survey of 635 students found that nearly one out of three would consider cheating in any environment and students indicated that they were more than four times as likely to cheat in an online class (Watson & Sottile, 2010). However, the same survey found no significant differences in student descriptions of cheating behavior in online and face-to-face classes (Watson & Sottile, 2010).

Many studies that address the prevalence of cheating on line vs. to face-to-face classes, many of these studies relied on student self-reports (Guyette & Piotrowski, 2009; Stuber-McEwen, Wisely, & Hoggatt, 2009; Etter, Cramer, & Finn, 2007; Watson & Sottile, 2010). Research focusing on actual student behavior has found conflicting results. For example, Ladyshewsky (2015) analyzed graduate student test scores and found no difference between the test scores in unproctored online tests when compared to face-to-face, proctored tests. Similarly, Yates and Beaudrie (2009) found no differences in course grades between community college students who took monitored versus unmonitored exams. Beck (2014) extended this work to examine scores on specific tests, where steps to reduce cheating such as randomizing the order of questions, having a single question on each page, and only allowing forward progress through the tests were used. Beck also found no differences between undergraduate student grades on monitored versus unmonitored tests (2014).

Other studies have found rampant cheating. For example, one large-scale study of cheating in online courses and work tasks found that between 26% and 34% of students cheated by looking up answers online, as did 20% of contract employees (Corrigan-Gibbs, Gupta, Northcutt, Cuttrell & Thiess, 2015). This innovative study used multiple techniques to identify cheating, including: 1) planting a fake resource that appeared in Google search engines when the

exact wording of the question was entered; 2) expert analysis of wording, comparing student responses to one another as well as to common website language focusing on idiosyncratic language; and 3) tracking of IP addresses. However, unlike a typical university class, both samples involved a degree of anonymity: the class was a massive open online course aimed at undergraduate engineering students in India, and the contract employees were identified and assigned the work through a crowdsourcing work platform.

In summary, when clear-cut differences in test scores occur in separate sections of the same course or when a test is taken under contrasting conditions, questions arise about potential underlying reasons for grade disparities. There are various strategies for addressing integrity during online tests, and the use of proctoring software is one of them (Berkey & Halfond, 2015).

Proctoring software involves two major elements. First, it activates the camera on a computer, and records the student taking the exam. This enables faculty to observe the students' behavior and identify activities that may indicate cheating such as talking to others or looking up information in books. Second, it either limits the students' ability to use their computers for other tasks by eliminating the ability to engage in activities such as copy-pasting, printing and searching the internet, or it records everything that students do on their computers, or both. Limiting students' abilities to use other tools or resources is referred to as "locking down" the computer or browser. Recordings of exams can be reviewed by the professor or teaching assistants; alternatively, they can be reviewed by employees of the proctoring vendor, either simultaneous to the exam or afterward, who mark points in the exam when possible violations of exam rules are identified.

The purpose of this study was to compare test performance of students enrolled in multiple sections of the same online class where four of the nine sections used proctoring software for at least one of their tests and the other five course sections never proctored tests. We also compared student scores in the same section with and without the use of proctoring software.

Methods

This study examined the effect of proctoring tests in an online undergraduate course, Medical Terminology (KNH 209), at Miami University, a public university located in southwestern Ohio with approximately 17,000 students. Medical Terminology is a lower level undergraduate elective class, with no pre-requisites. All university students enrolled as full or part time students are eligible to take the class. It satisfies requirements toward graduation in virtually all academic divisions. Twenty students enrolled in each of nine sections of this course, totaling 180 undergraduates with the following majors: accountancy, athletic training, biochemistry, biology, economics, English, finance, public health, media studies, kinesiology, mechanical engineering, microbiology, nutrition, political science, psychology, Spanish, and speech pathology/audiology, sport leadership and management, communication, supply chain management, and zoology. All nine instructors agreed to use common exam formats that apply concepts from WCET's best practice for online education, including timed tests, random questions from a common question pool, and responses that are in randomized order (WCET, 2009).

Of the nine sections of this course, four used proctoring software. Three instructors selected a few of the tests to be proctored using Software Secure (http://www.softwaresecure.com/), a remote proctoring software that videotapes the student in their surroundings, blocks some unauthorized activities on the computer, and records students' desktops during the test. Software Secure uses live proctors, who review the recordings after the exam and identify likely situations of cheating. Two proctors, certified by the vendor, review every test. The tool also requires students to scan the room in which they are taking their exam. One instructor had all of the tests proctored using Respondus Monitor (http://www.respondus.com/products/monitor/index.shtml), which utilizes both locking down the browser and videotaping the student taking the test.

Following the completion of the tests, videos from Software Secure were reviewed by the company to detect rule violations or suspicious activity. The instructor for the course received feedback of the review and could watch the videos at each point of a potential breach to confirm if a violation occurred. Respondus Monitor generates a set of thumbnails of the full video recording that can be reviewed and flagged by the instructor for potential violations. The instructor can click on each thumbnail to watch that segment of the full video recording of the student taking the quiz. Five instructors did not use proctoring software options, while one instructor used only Lockdown Browser (no video recording or review) for half of the tests.

Students in all nine sections were informed that tests were to be taken by themselves with no notes or other resources allowed during the test. Students in the sections that were proctored were not certain of the exact test(s) throughout the course that would be proctored prior to the start of the test. Tests varied in terms of time limits, number of questions, and proctoring, but all covered similar material, and questions were randomly drawn from a shared question bank. Table 1 provides a summary of the nine class sections and indicates which of the tests in each section were proctored.

Table 1
Four Quiz Conditions (P=proctored, U=unproctored, U/L=unproctored/lockdown) for KNH 209/Medical Terminology sections A through I in January 2015

Quiz	Section A	Section B	Section C	Section D	Section E	Section F	Section G	Section H	Section I
1	U	U	P	U	U	U	U	U	U
2	U	P	P	U	U	U	U	P	P
3	U	U	P	U/L	U	U	U	U	U
4	U	P	P	U/L	U	U	U	U	U

U: Unproctored
U/L: Unproctored and lockdown only, no video monitoring
P: Proctored with video monitoring (Software Secure or Respondus Monitor)

Table 2 reports the number of students who were proctored or unproctored on each quiz. Student enrollments were tracked in all sections. Following the conclusion of the course, all students were contacted about the use of their data in class with all identifiers removed, and were provided an opportunity to have their data omitted from analyses.

Table 2
Total number of quizzes that were Proctored and Unproctored in 9 sections of KNH 209/Medical Terminology in January 2015

Quiz 1	One quiz (n=14) was proctored; eight quizzes (n=148) were unproctored
Quiz 2	Three quizzes (n=48) were proctored; six quizzes (n=109) were unproctored
Quiz 3	Two quizzes (n=31) were proctored; seven quizzes (n=129) were unproctored
Quiz 4	Two quizzes (32) were proctored; seven quizzes (n=130) were unproctored

Of the initial 180 students enrolled, 22 dropped the course. Of the 158 students who completed the course, 11 did not complete all tests. The anonymized data from the 147 students who consented and had completed all four tests were then used in a statistical analysis to assess the effect of proctoring on test scores and percentage of allotted time taken.

Data Analysis

The impact of proctoring on student quiz performance was evaluated using a linear mixed effects model (Verbeke & Molenberghs, 1997; Montgomery, 2013). A realistic assumption can be made that responses from tests taken by the same student or students with the same instructor may be related; thus linear mixed effects models are used to allow for these relationships to be reflected in the correlation structure of our analysis. First we aimed to model the test score percentages to assess the effect of proctoring. Due to a concern that the difficulty of the four exams may not be uniform in difficulty of material and that the number of questions per test may have effect on test scores, we consider these as covariates in the modeling. The model selection based on the Bayesian Information Criterion (BIC) confirms the importance of accounting for these factors. The selection process yielded a model with fixed effects for tests, proctoring administration and number of questions on the test, and random effects for sections and for students within sections. The linear mixed effects model for test score percentage was parameterized as:

Model Equation 1.

$$Score_{ijk} = Test_k + \beta_L I_{ijk}(Lockdown) + \beta_P I_{ijk}(Proctored) + \beta_Q NumQ_{ik} + \delta_i + \gamma_{ij} + \epsilon_{ijk}$$

where we model the score of the kth test for the jth student in the ith section using a cell means parameterization of test averages, μ_k, which use non-proctored exams with 20 questions as the baseline. The model terms associated with the fixed effects are defined as:

$Test_k$ Average score on test k with no proctoring software and 20 questions (baseline)

β_L Additive change to baseline score when Lockdown (no video) used on test

β_P Additive change to baseline score when video proctoring used on test

$I_{ijk}(.)$ Indicator function for use of proctoring software in test k for student j in section i

β_Q Additive change to baseline score for every additional questions above the baseline

$NumQ_{ik}$ The number of questions beyond than the baseline of 20 on test k in section i.

The model terms i(section), ij(student) and εijk (error) are nested random effects that are specified such that:

$$Cov(Score_{ijk}, Score_{lmn}) = \begin{cases} 0, & \text{if sections } i \neq l \text{, student } j \neq m \text{, and test } k \neq n \\ \sigma^2, & \text{if sections } i \neq l \text{, student } j \neq m \text{, and test } k = n \\ \sigma^2 + \sigma_\gamma^2, & \text{if sections } i \neq l \text{, student } j = m \text{, and test } k = n \\ \sigma^2 + \sigma_\gamma^2 + \sigma_\delta^2, & \text{if sections } i = l \text{, student } j = m \text{, and test } k = n \end{cases}$$

It was also speculated that academic dishonesty on online tests may manifest as longer times taken on the tests due to the extra time spent searching through prohibited reference materials. To explore the impact of proctoring software on the time taken to complete the tests we fit a linear mixed effects model to the percentage of allotted time used. Note that the metric used in modeling differences in time usage was the percentage of allotted time used by the student; this is to maintain a consistent interpretation with different numbers of questions and time allowed across the sections. Model selection and diagnostics were run in the same fashion as in the model for test scores, and the model covariates and random effects for the selected model turn out to follow an identical structure to those in Equation (1) above. The model for percentage of time taken follows the form,

Model Equation 2.
$$Time_{ijk} = Test_k + \beta_L I_{ijk}(Lockdown) + \beta_P I_{ijk}(Proctored) + \beta_Q NumQ_{ik} + \delta_i + \gamma_{ij} + \epsilon_{ijk} .$$
(2)

Data cleaning, data summaries, visual graphics and linear mixed models and diagnostic tools were created using the R software using the dplyr (Wickham & Francois, 2015), ggplot2 (Wickham, 2009) and nlme (Pinheiro, Bates, DebRoy, Sarkar & R Core Team, 2015) packages.

Results

Figure 1 visually presents the scores and times taken on tests within each class section and is colored to emphasize the proctoring status of each test group. A test was considered proctored when it included videotaping. We see that there are noticeable differences in proctored and unproctored exams, primarily that proctored exams seem to have lower scores and take a larger percentage of the allotted time. The average test scores for proctored tests was 74.3% (SD=12.3) and 89.4% (SD=9.0) for unproctored tests. The average percentage of allotted time taken on proctored tests was 20.4% (SD=13.9) and unproctored tests was 41.2% (SD=14.1); showing that students took approximately half the amount of time taking proctored test compared with unproctored tests. Note that unproctored tests with lockdown only (no video monitor) had an average score of 93.2% (SD = 5.9) and took an average of 40.0% (SD=10.1) of the time allotted; quite comparable in test scores and time used with the unproctored tests. See Table 3 for a full listing of statistics for test scores and percent of allotted time used within proctoring groups.
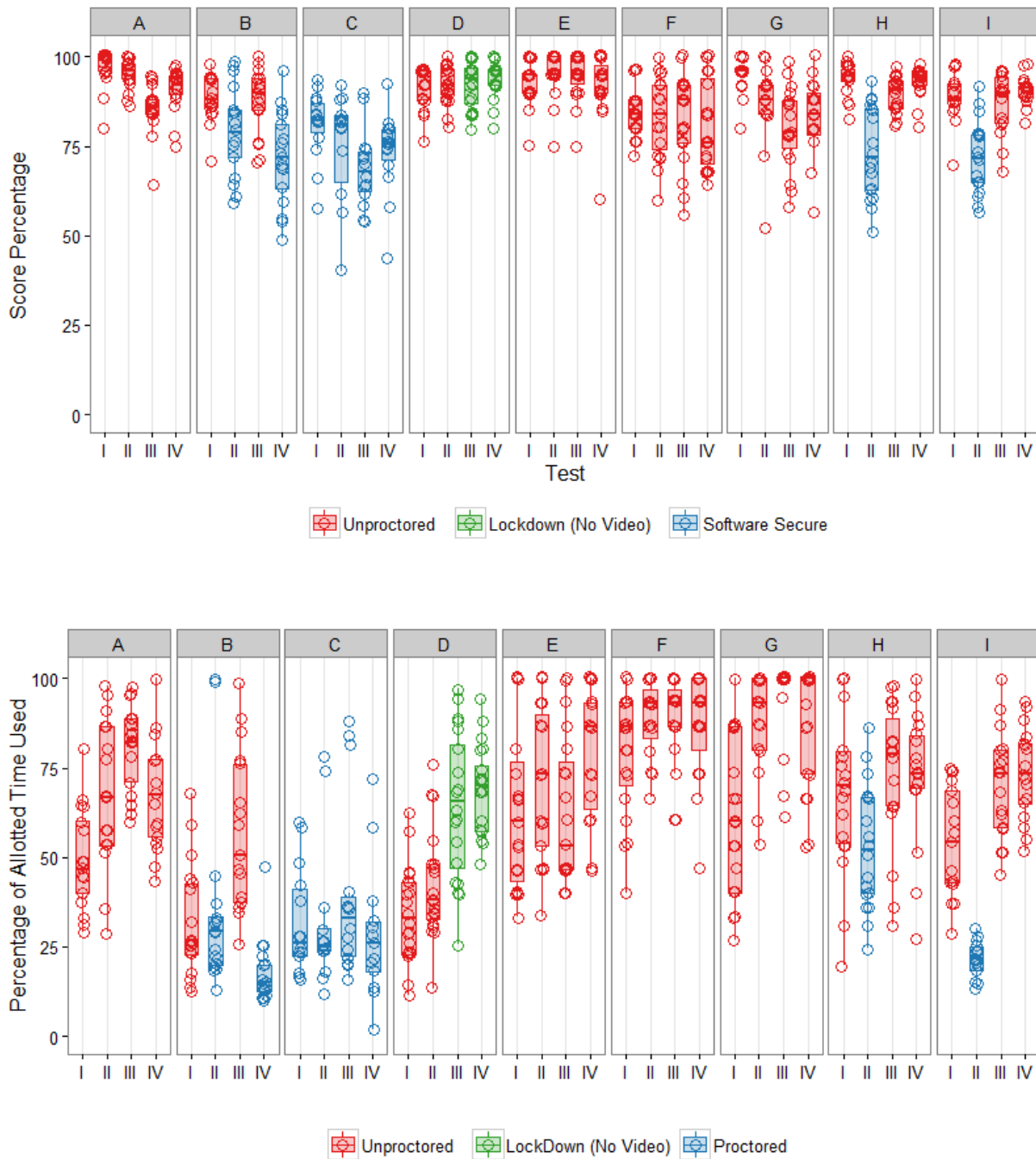
Figure 1. Test scores (%) and time used (% of allotted) in nine Sections (A-I), colored by proctoring status. Proctored tests (Blue) tended to score lower and take less time than unproctored tests (Red). Tests with Lockdown (Green) behaves similar to unproctored sections.

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)

Table 3
Proctor Status and Average Test scores, Percent Time Used, and Number of Total Tests and Students

Proctor status	Average + [SD] Test Score (% correct)	Average + [SD] Percent Time Used (% of time given)	Number of Tests	Number of Students
Unproctored	89.4 [12.3]	41.2 [14.1]	471	147
Proctored *with* video monitor	74.3 [5.9]	20.4 [13.9]	125	66
Lockdown (no video monitor)	93.2 [9.0]	40.0 [10.1]	40	20

We turn to the fitted models discussed in the analysis section above to assess the significance of the proctoring related difference seen in the visual and numerical exploration. Table 4 shows the summary of the effects of proctoring on test scores, as estimated from the linear mixed effects model. There did not appear to be any trend or extreme outliers in the residuals, hence, the use of this model appears to be justified. The baseline means for tests 1 through 4, unproctored tests with 20 questions, were: 89.7, 87.8, 83.4, and 84.8, respectively. This accounts for general differences in difficulty, where the first two tests were less difficult than the last two tests. The differences in the test scores in the model are statistically significant ($p<0.05$). Tests proctored with the Software Secure video monitoring were found to have significantly lower test scores than unproctored test scores. The video proctored tests were found to score 17.2 percentage points (95% CI: [4.8, 19.6]) *lower* than unproctored tests. This implies a significant, and substantial, decrease in scores under video proctoring, after controlling for differences in test difficulty and number of questions. Tests that used only Lockdown (no video) were found to have a score 7.4 percentage points (95% CI: [3.9, 11.2]) *higher* than unproctored tests, after controlling for differences in test difficulty and number of questions. While this implies students taking a test using only Lockdown (no video) have a significant improvement in scores, there is only one section that implemented this technology; thus confounding the effect of lockdown and instructor.

Table 4
Fitted Coefficients and Variance Estimates for Linear Mixed Effects Model for Test Score Percentages, as Parameterized in Model Equation

Fixed Effect	Model Term	Estimate	95% CI
Test 1 baseline (unproctored/20 questions)	$Test_1$	89.77	(86.86 , 92.68)
Test 2 baseline (unproctored/20 questions)	$Test_2$	87.80	(84.46 , 91.14)
Test 3 baseline (unproctored/20 questions)	$Test_3$	83.37	(80.12 , 86.62)
Test 4 baseline (unproctored/20 questions)	$Test_4$	84.76	(81.32 , 88.20)
Lockdown (no video) effect	β_L	7.54	(3.92 , 11.15)

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)

Proctored (Software Secure or Respondus Monitor) effect	β_P	-17.23	(-19.62 , -14.83)
Additional questions effect	β_Q	0.13	(0.05 , 0.21)

Random Effect	Variance Term	Variance Estimate	Percentage of Total Variance
Section	σ_δ^2	10.1	11.2 %
Student	σ_γ^2	28.3	31.3 %
Residual error	σ^2	52.0	57.5 %

Not only did proctoring of tests affect test scores but there was evidence that proctoring also affected how long students took to finish. Table 3 shows that when students were unproctored, including only Lockdown (no video), they used much more of their available time than if the test was proctored with video. Table 5 contains the effects from the linear mixed effects model for the percentage of allotted time taken. The baseline tests, with 20 unproctored questions, show that students tended to take more time to complete the later exams. There was no significant effect of number of questions on the percentage of allotted time taken, indicating that the time per questions was sufficiently similar to allow comparison across sections. Lastly, the proctored group took an estimated 30.5 percent less of the time allotted (95% CI: [25.4 , 35.7]) in completing their exams than the unproctored students.

Table 5
Fitted Coefficients and Variance Estimates for Linear Mixed Effects Model for Percentage of Allotted Time Taken on Tests, as Parameterized in Model Equation 2

Fixed Effect	Model Term	Estimate	95% CI
Test 1 baseline (unproctored/20 questions)	$Test_1$	56.10	(46.05 , 66.16)
Test 2 baseline (unproctored/20 questions)	$Test_2$	69.67	(58.87 , 80.48)
Test 3 baseline (unproctored/20 questions)	$Test_3$	71.43	(60.80 , 82.06)
Test 4 baseline (unproctored/20 questions)	$Test_4$	70.50	(59.50 , 81.49)
Lockdown (no video) effect	β_L	19.75	(11.95 , 27.57)
Proctored (Software Secure or Respondus Monitor) effect	β_P	-30.53	(-35.69 , -25.36)
Additional questions effect	β_Q	-0.01	(-0.20 , 0.18)

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)

Random Effect	Variance Term	Variance Estimate	Percentage of Total Variance
Section	σ_δ^2	184.51	34.2 %
Student	σ_γ^2	130.33	24.2 %
Residual Error	σ^2	224.17	41.6 %

The results of the linear mixed effects models for test scores and percentage of allotted time used show us that unproctored tests had significantly higher scores and took significantly more time than proctored test, while controlling for test ordering and number of questions. We see the dramatic difference in testing behavior in Figure 2 which contains the scatterplot of test scores and percentage of allotted time taken, colored by proctoring status. These finding are consistent with the suspicion that academic dishonesty, in the form of students searching through prohibited reference materials during the test, is more prevalent on unproctored exams.
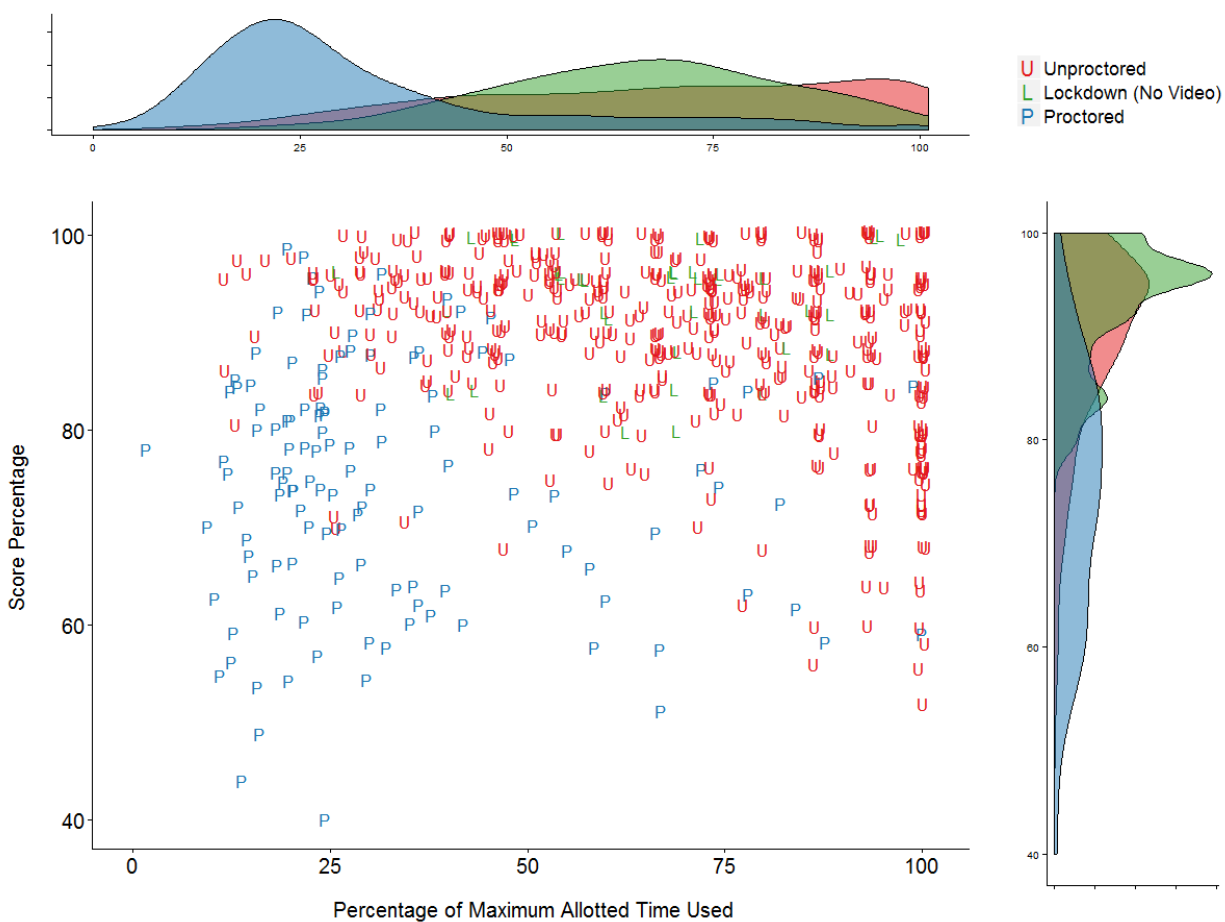

Figure 2. Plot of test score (%) vs. amount of time used (% max) for all sections combined. Points correspond to students in sections that were proctored using Software Secure or Respondus Monitor with Lockdown (Blue P), used Lockdown alone (Green L) or unproctored

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)

(Red U). Once again we see that the tests behaviors are similar for lockdown and unproctored tests.

Discussion

These results indicate clear and significant grade disparities in comparing test scores when the students took online tests that were proctored with video monitoring versus unproctored or unproctored with Lockdown (no video). Proctored test (with video) scores are significantly lower than unproctored test scores. The model fit also shows the same results as the difference between proctored and unproctored test scores is between 14 and 20 points, the difference of one or two letter grades. This difference occurred in students between multiple sections of the same course as well as within sections when the same students took tests proctored versus unproctored.

Test scores are not the only component factoring into student grades, as forum postings, case studies, homework assignments, blogs, and other types of work all contributed to the final grade in this course. Nevertheless, the striking difference in scores from proctored versus unproctored tests appeared to factor significantly into final grades as evidenced by the different final grade distributions. Sixty three percent of all students in sections with only unproctored tests earned an A, whereas 17% of all students in sections with proctored tests earned an A.

Another concern is the difference in attrition between the sections that offered proctored versus unproctored tests. Only seven of the 100 students initially enrolled in sections with unproctored tests dropped the class compared with 15 of 80 students initially enrolled in sections with proctored tests who dropped. Although no inquiries were made as to why students dropped the class, more than twice as many students in the proctored group dropped compared with the unproctored group.

Bunk, Li, Smidt, Bidetti, and Malize (2015) explored faculty perceptions in explaining negative attitudes toward online classes. While proctoring and academic honesty were not directly mentioned, faculty did express concern about compromised educational quality in online classes. In a study on student and faculty views of academic dishonesty and online learning, both faculty and students agreed that it would be easier to cheat in online classes (Kennedy, Nowak, Raghuraman, Thomas, & Davis, 2000). Methods suggested by faculty to counter cheating, included supervised final exams counting for a high percentage of the course grade, changing assignments each semester, using personalized assignments, verification software, and using open-book exams. Proctoring software was not mentioned by faculty, because it was not commonly available at the time.

A study by Spaulding (2009) did not provide compelling evidence for an increased prevalence of academic dishonesty in online vs. traditional classes, which may lead many faculty to underestimate the frequency of academic dishonesty in their classes. Given this perception, Hard, Conway, and Moran (2006) reported that faculty members who perceive academic dishonesty as rare do not actively work to prevent it. Investigating academic integrity is complicated, whether in traditional or face-to-face learning and testing environments. Student perception of cheating online may be different than in a face-to-face situation (Rains, et al.,

2011) even when instructions clearly state otherwise. The potential for academic dishonesty (Corrigan-Gibbs et al., 2015; Jones, Blankenship, & Hollier, 2013; Moten, et al., 2013) and the perception that cheating occurs more frequently in online classes (Grijalva, Nowell, & Kerkvliet, 2006; Raines et al, 2011) present challenges to all stakeholders. Much research argues that cheating is prevalent in online courses, but few studies measure actual cheating behavior. Some found evidence of significant cheating in online tests (Corrigan-Gibbs, et al., 2015), while others did not (Ladyshewsky, 2015). The current study did not assess cheating behavior. Instead, it compared test scores when students used proctoring software with those that were unproctored. Disparate test grades imply that cheating likely occurred when student tests were unproctored, especially given the large and statistically significant grade difference of 17 points, representing an average difference of two letter grades between scores on tests when proctoring software was used versus when it was not.

This study provides substantive evidence of disparate test results in online courses, as indicated by significantly higher scores both within classes and across class sections on unproctored versus proctored online tests. After controlling for the effects of test difficulty and student and teacher differences, students taking proctored online exams scored approximately 17 points lower out of 100 when compared to unproctored students. The different scores approximated an average test grade of A to A- on unproctored tests and C to C- on proctored tests. Furthermore, when unproctored, students took significantly more time to complete tests. It is possible that students used the extra time to look up answers, despite the application of testing best practices of providing limited time, randomized selection of items, and instructions stating that using resources during a test was not allowed.

This potential for academic dishonesty cannot be ignored (Harbin & Humphrey, 2013). Previous research on student perceptions about whether they felt they might cheat in online versus face-to-face test conditions have been inconsistent, however, it appears that in this current study, the finding by Watson and Sottile (2010), where students indicated they would be more than four times likely to cheat in an online class, seemed to have occurred, with the grade distribution indicating that students taking unproctored online tests were four times more likely to receive a grade of A compared with students who took proctored online tests. Concerns about the integrity of online courses due to cheating and fraud have reached the popular press (Newton, 2015). There are real consequences for students who cheat, who may not learn critical content for thinking, problem solving, and foundational information required for upper level course work. Additionally, the reputation of faculty and institutions and student learning are compromised when acts of cheating are not addressed. Faculty and institutions will need to confront the likelihood that breaches in academic honesty occur in all class formats. In online classes, in particular, proactive interventions that include proctoring software with video monitoring may deter cheating and protect academic integrity.

Limitations of this Study

It is important to consider potential limitations to the generalizability of these results. This was a class of medical terminology, which requires that technical terms be memorized and accurately applied, and where assessment included multiple choice tests. It is not clear that the size of the effect would be as large with courses that do not involve timed, closed-ended tests. In

addition, these were classes populated by traditional students in a Midwestern, public university. Again, the potential for generalizability to other populations may be low.

Conclusions and Future Studies

Students enrolled in online courses in which at least one online test was proctored with video monitoring scored on average 17 points lower than students enrolled in the same courses with no test proctoring. The effect of *proctoring with video* is large enough to suggest that an impact on test scores exists, with the likelihood that when unproctored, students may resort to academic dishonesty by using resources that were explicitly forbidden during the test. The effect of proctoring with video shows a potential effect on the percentage of test time used to take the test, with proctoring resulting in less time compared with unproctored tests, where students took significantly more time to complete the test. Additionally, lockdown software without video monitoring, did not have a similar impact as proctoring software that used video monitoring. Proctoring with video monitoring significantly negatively impacts online test grades, probably because it deters cheating, and its use is important to assure academic integrity through similar test taking conditions in similar courses when using online tests.

It would be interesting to replicate this study or use a randomized design in other courses and at other universities. In addition, the different proctoring tools themselves could be examined. As online test proctoring becomes more common, faculty and students may learn about advantages and disadvantages of different vendors and systems. For example, it may be fruitful to examine possible differences between vendors that employ human proctors as opposed to fully-automated proctoring systems. While future research may affect the proctoring choices, these results point to the need for proctoring software to contribute to the integrity of online testing.

Correspondence concerning this article should be addressed to Beth Rubin, Assistant Provost for eLearning, Miami University, Oxford Ohio 45056. Contact: rubinb@miamioh.edu

References

Allen, I.E., and Seaman, J. (2015). Grade level: Tracking online education in the United States. Babson Survey Research Group and Quahog Research Group, LLC. Last access on February 3, 2016: http://www.onlinelearningsurvey.com

Beck, V. (2014). Testing a model to predict online cheating: Much ado about nothing. *Active Learning in Higher Education, 15(1)*, 65-75.

Bunk, J., Li, R., Smidt, E., Bidetti, C., and Malize, B. (2015. Understanding faculty attitudes about distance education: The importance of excitement and fear. *Online Learning, 19(4), 1-11*.

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

Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., and Thies, W. (2015). Deterring cheating in online environments. *ACM Transactions on Computer-Human Interaction, 22*(6), Article 28. Last access on February 3, 2016: DOI: http://dx.doi.org/10.1145/2810239

Etter, S., Cramer, J.J., and Finn, S. (2007). Origins of academic dishonesty: Ethical orientations and personality factors associated with attitudes about cheating with information technology. *Journal of Research on Technology in Education*, *39(2),* 133-155.

Grijalva, T.C., Nowell, C., and Kerkvliet, J. (2006). Academic honesty and online courses. *College Student Journal*, *27(3)*, 180-185.

Harbin, J. L., and Humphrey, P. (2013). Online cheating: The case of the emperor's clothing, elephant in the room, and the 800 lb. gorilla. *Journal of Academic and Business Ethics*, *7*, 1-6.

Hard, S.F., Conway, J., and Moran, A.C. (2006). Faculty and college student beliefs about the frequency of student academic misconduct. *The Journal of Higher Education, 77(6), 1058-1080.*

Jones, I.S., Blankenship, D., and Hollier, G. (2013). Am I cheating? An analysis of online students' perceptions of their behaviors and attitudes. *Psychology Research, 3*(5), 261-269.

Kennedy, K., Nowak, S., Raghuraman, R., Thomas, J., Davis, S.F. (2000). *College Student Journal, 34(2), 309*.

King, C.G., Guyette, R. W., & Piotrowski, C. (2009). Online exams and cheating: An empirical analysis of business students' views. *The Journal of Educators Online, 6, 1*, 1-11.

Ladyshewsky, R.K. (2015). Post-graduate student performance in 'supervised in-class' versus 'unsupervised online' multiple choice tests: implications for cheating and test security. *Assessment and Evaluation in Higher Education*, 40(7), 883-897. DOI: 10.1080/02602938.2014.956683

Montgomery, D. (2013). Experiments with Random Factors. In *Design and Analysis of Experiments* (8th ed.). New York: Wiley.

Moten Jr., J., Fitterer, A., Brazier, E., Leonard, J., and Brown, A. (2013). Examining online college cyber cheating methods and prevention methods. *The Electronic Journal of eLearning, 11, (2)*, 139-146.

Newton, D. (2015). Cheating in online classes is now big business. *The Atlantic, 4*. Retrieved from http://www.theatlantic.com/education/archive/2015/11/cheating-through-online-courses/413770

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2015). *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1. http://CRAN.R-project.org/package=nlme

Raines, D.A., Ricci, P., Brown, S.L., Eggenberger, T., Hindle, T., and Schiff, M. (2011). Cheating in online courses: The student definition. *The Journal of Effective Teaching, 11*(1), 80-89.

SAS Institute Inc. 2012. *Base SAS® 9.4* Cary, NC: SAS Institute Inc.
Spaulding, M. 2009. Perceptions of academic honesty in online vs. face-to-face classrooms. *Journal of Interactive Online Learning, 8(3), 183-198.*

Stuber-McEwen, D., Wisely, P., and Hoggatt, S. (2009). Point, click, and cheat: Frequency and type of academic dishonesty in the virtual classroom. *Online Journal of Distance Learning Administration,* 12(2). Retreived from http://www.westga.edu/~distance/ojdla/fall123/stuber123.html

Verbeke, G. and Molenbergh, G. (1997). *Linear Mixed Models in Practice: A SAS-Oriented Approach.* New York, NY: Springer.

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

WCET (2009). *Best practice strategies to promote academic integrity in online education.* WCET, UT TeleCampus, and Instructional Technology Council. Retrieved from http://wcet.wiche.edu/sites/default/files/docs/resources/Best-Practices-Promote-Academic-Integrity-2009.pdf

Wickham,H, and Francois, R. (2015). *dplyr: A grammar of data manipulation*. R package version 0.4.3. Retrieved from. http://CRAN.R-project.org/package=dplyr

Wickham, H. (2009) *ggplot2: Elegant graphics for data analysis.* New York, NY: Springer.

Yates, R.W., and Beaudrie, B. (2009). The impact of online assessment on grades in community college distance education mathematics courses. *American Journal of Distance Education, 23*, 62–70.