

November – 2018

Interaction of Proctoring and Student Major on Online Test Performance



Helaine M. Alessio¹, Nancy Malay¹, Karsten Maurer¹, A. John Bailer¹, and Beth Rubin²
¹Miami University, Oxford Ohio, ²Campbell University, Buies Creek, NC

Abstract

Traditional and online university courses share expectations for quality content and rigor. Student and faculty concerns about compromised academic integrity and actual instances of academic dishonesty in assessments, especially with online testing, are increasingly troublesome. Recent research suggests that in the absence of proctoring, the time taken to complete an exam increases significantly and online test results are inflated. This study uses a randomized design in seven sections of an online course to examine test scores from 97 students and time taken to complete online tests with and without proctoring software, controlling for exam difficulty, course design, instructor effects, and student majors. Results from fixed effects estimated from a fitted statistical model showed a significant advantage in quiz performance (7-9 points on a 100 point quiz) when students were not proctored, with all other variables statistically accounted for. Larger grade disparities and longer testing times were observed on the most difficult quizzes, and with factors that reflected the perception of high stakes of the quiz grades. Overall, use of proctoring software resulted in lower quiz scores, shorter quiz taking times, and less variation in quiz performance across exams, implying greater compliance with academic integrity compared with when quizzes were taken without proctoring software.

Keywords: online learning, online testing, academic integrity, academic honesty, proctoring, distance learning

Introduction

The dissemination of advanced technology in education, and in particular the growth of online and distance learning courses, have created countless opportunities for intellectual and professional growth. Prospects for continuing one's education without traditional constraints of in-class instruction schedules appeal to many learners, especially those who can learn without being under direct in-person supervision of an instructor. An unintended consequence, however, is an increased potential for academically dishonest behaviors due to opportunities for cheating that use unauthorized technological assistance and occur out of the sight of an instructor (Etter, Kramer, & Finn, 2006). This is a significant problem in higher education as academic integrity is critical to an institution's reputation, as well as the expectation of workplaces and society that college graduates actually master the content and skills assessed in their program of study. Despite efforts to encourage honesty in all types of course assessments, higher education institutions face the same types of scandals and deceit that occur in the workplace and society (Boehm, Justice, & Weeks, 2009).

Academic misconduct has many forms that include submitting work that is not one's own, plagiarizing other's words without acknowledgement, using unauthorized notes during an exam, receiving help from another person or from the internet during an exam, programming answers into electronic devices, texting answers, and having another person take an exam or write a paper in one's place. Newly established internet business sites sell or trade academic papers and answers to tests in specific courses at colleges and universities across the nation (Berkey & Halfond, 2015). The convenience and perceived anonymity associated with patronizing these sites can make it difficult to resist when students perceive that the stakes are high if they do not achieve a high grade. The numerous cheating methods, especially those that use the latest technology, make it difficult for even experienced instructors to detect.

Research on student perceptions about integrity indicates that student dishonesty is a significant concern, especially in online classrooms (Berkey & Halfond, 2015; D'Souza & Siegfeldt, 2017; Kitahara, Westfall, & Mankelwicz, 2011), and that steps taken to ensure a fair environment when it comes to assessment are supported by students as well as instructors. Faculty who teach online are encouraged to use various pedagogical strategies to develop a relationship of trust with their students (WCET, UT Telecampus, & Instructional Technology Council (WCET), 2009). Connecting with students in meaningful ways is important, but can still be challenging due to students who are geographically dispersed. Student-instructor relations can be further challenged when teaching and assessment roles become separated due to automation in popular modular teaching systems (Amigud, Arnedo-Moreno, Daradoumis, & Guerrero-Roldan, 2017). The sense of distance, weak personal ties to classmates and instructors, and perceived anonymity, may yield a detached feeling that enables a student to engage in dishonest behavior in an online assessment (Corrigan-Gibbs, Gupta, Northcutt, Cutrell, & Thies, 2015). Nevertheless, the success of distance learning requires careful attention to the design of the course as WCET (2009) describes, including establishing policy, incentives for honesty, and holding accountable students who demonstrate dishonesty.

Background

Academic Dishonesty in Online Classes

Cheating, while not new to academia, has become increasingly complex in online environments where asynchronous learning and assessment occur far from the instructor's explicit monitoring. Students may be tempted to cheat due to the perception that academic dishonesty will go unnoticed in a virtual classroom. Instructors are challenged with providing an environment and tools that prevent and detect occurrences of academic dishonesty. There is a growing body of literature supporting the notion that students are more likely to cheat in online courses than in face-to-face environments. This includes both indirect and direct measures of cheating in a wide variety of educational contexts.

Self-report survey research includes a study by King, Guyette, and Piotrowski (2009), who found that 73% of 121 undergraduate students reported that it was easier to cheat online compared to a traditional face-to-face classroom. Furthermore, Watson and Sottile (2010) reported that when asked if they were likely to cheat, 635 students surveyed indicated they were more than four times as likely to cheat in an online class than in a face-to-face class.

Direct assessment of cheating has also found it to be common in online courses. Corrigan-Gibbs and his colleagues (2015) directly measured cheating in online MOOCs and work assignments, using both content analysis of open-ended assignments and visits to a "honey pot," a website that promised solutions to problems; they found that between 13% and 34% of students cheated, despite honor codes and warnings of penalties. Fask, Englander, and Wang (2014) found evidence that students taking an online unproctored test cheated more than those taking the same test in a proctored face-to-face format. Despite a few studies that found no evidence of cheating (e.g., Beck, 2014), these results present an ominous picture of integrity in the online classroom that Rujoiu and Rujoiu (2014) reported is associated with integrity or lack thereof, in the workplace.

Factors That Affect Cheating

Understanding factors that influence student behaviors to cheat is complex as it includes personal factors and ethical principles, regardless whether cheating behavior occurs in a traditional classroom or in technologically-assisted ones such as online classes (Etter et al., 2006; McCabe, Trevino, & Butterfield, 2001). Personal factors can include individual situations and circumstances, including each student's prior experiences, level of competence, and beliefs, that guide their behaviors in the classroom (Schuhmann, Burrus, Barber, Graham, & Elikai, 2012). Ethical principles can be influenced by personality and peers, as well as the organizational climate, condition, and structure of the classroom.

The classroom environment that is created by the instructor is important in affecting student behaviors of all types: frequency and quality of participation in class, workgroup cooperation, sparking student curiosity, independent learning, and demonstrating academic integrity on assignments and assessments. Rubin and Fernandes (2013) summarized several reports on organizational climate and composition theory and found evidence that the psychological climate in online classes facilitates students' interpretation and affects the action they take, which in turn affects the environment, continuing in a reciprocal way.

D'Souza and Siegfeldt (2017) describe the academic dishonesty triangle of three factors that contribute to cheating: "Incentive to cheat, an opportunity to cheat, and rationalization to cheat" (p. 274). According to this framework, taking an unproctored online test provides an opportunity to cheat. Taking majors that lead to highly competitive graduate education, such as medical, law, or business schools, or that require high grades to maintain student status, would constitute an incentive to cheat. If this model is correct, such students would be more likely to cheat than would students in less competitive majors.

Another factor that may create an incentive to cheat is the perceived difficulty of a test (Christie, 2003). If students believe that they will not likely be successful without cheating, or that their academic success rests upon their performance on an exam such as occurs in high-stakes testing, it gives them a greater incentive to cheat. Students who have higher cumulative GPAs are less likely to cheat than are students with lower GPAs, indicating different levels of preparation (Schuhmann et al., 2012). Students taking courses within their major may be less likely to cheat due to greater interest or preparation.

Multiple studies have addressed ways of reducing the likelihood of cheating on online assessments. Strategies that emerged from studies by Beck (2014), D'Souza and Siegfeldt (2017) include various aspects of test and course design such as offering multiple versions of tests or even randomly selecting questions from a pool; providing a tight testing time-limit; randomizing questions and options; reducing closed-ended assessment to reduce the stakes of testing; blocking students from printing the exam questions; withholding answers until the exam is completed by all students; avoiding high-stakes tests; and developing a supportive and trusting community within the class (Beck, 2014; D'Souza & Siegfeldt, 2017; McCabe et al., 2001; Rogers, 2006; WCET, 2009). These techniques have been assessed in combination rather than separately in an experimental format, so it is not yet possible to know which approaches have been more effective (e.g., Beck, 2014; Cluskey, Ehlen & Raiborn, 2011; McGee, 2013),

Some researchers hold that appropriate instructional design of open courses can eliminate cheating, particularly when assessment relies upon application of concepts rather than memorization of facts (Cluskey et al., 2011; McGee, 2013). However, several studies belie this notion. Northcutt, Ho, and Chuang (2016) found that a significant number of students taking Massive Open Online Courses (MOOCs) cheat by means of using more than one user account: one to "harvest" questions and correct answers, and another to obtain a certificate. This large-scale study used multiple algorithms to identify such cheating, and found 657 individuals across 115 courses used at least one cheating strategy called Copying Answers Using Multiple Existences Online (CAMEO). Among those students who earned 20 or more MOOC certificates, 25% appeared to use the CAMEO method to cheat. In another study of large-scale open online courses (Corrigan-Gibbs et al., 2015), assessment involved the application of concepts and high levels of critical thinking in both closed and open-ended questions, whereby instructors created 15 versions of the test with randomized questions arbitrarily pulled from a question bank to each student in the course. Despite these aspects of assessment design, the researchers found that a large proportion of students (13% - 35%) cheated by sharing answers with other students, seeking correct answers online, or using the CAMEO method.

Prior Findings of Grade Disparities and Cheating in Online Classes

Previous research has used the finding of significant differences in scores between proctored and unproctored tests as a measure of cheating (Beck, 2014; D'Souza & Siegfeldt, 2017). An elegant study by

Fask et al. (2014) statistically controlled for the effects of online versus face-to-face examination processes, and found that students were more likely to cheat in unproctored online tests. A study by Alessio et al. (2017) attempted to determine if online quiz results were lower when proctored than when unproctored, which would imply, although not directly prove, that cheating occurred more often in online quizzes that were not proctored. In a natural design study of 147 students enrolled in nine sections of the same online course, student scores averaged 17 points higher when they were not proctored compared to when they were proctored. This result was consistent both within and between sections. Students who were not proctored also used significantly more time to complete their online quizzes compared to those who were proctored, a finding both within and between sections. This finding appears to support students' attitudes toward cheating in an online class as reported by Watson and Sottile (2010), but also suggests an intervention strategy likely to prevent cheating - the use of online proctoring software.

The study described in this manuscript uses a randomized design in multiple sections of an online course to examine quiz scores and time taken to complete online quizzes with and without proctoring software, controlling for exam difficulty, course design, instructor effects, and student majors, in an effort to explore some attributes that may affect academic dishonesty. The research questions of this randomized study are:

1. What is the significance of proctoring software on student's test scores and completion time in comparison to tests without proctoring?
2. What is the effect of exam difficulty on scores in online tests with and without the use of proctoring software?
3. What is the effect of student major on scores in online tests with and without the use of proctoring software?

Methods

Investigating Academic Integrity in an Online Course

The data were collected from college students attending a college in the Midwest region of the United States taking an accelerated format, three-week course titled *Medical Terminology for Health Professionals*. The bulletin description states that this course “provides the opportunity for students to comprehend basic terms related to anatomy, pathophysiology, diagnostics and treatment. Students will understand word parts necessary to build medical terms and acceptable medical abbreviations and symbols.” The course is a common prerequisite for professional schools in many allied health fields, including medical, physical therapy, nursing, and occupational therapy schools. The class emphasizes the learning and application of medical vocabulary terms associated with anatomy, health, and disease. Following best practices for reducing cheating, the course includes multiple forms of assessment, including open-ended discussions (e.g., case studies that require accurate descriptions of medical conditions, problems that require use of commonly used and standard medical terms, and creation of subjective, objective, assessment, and plan [SOAP] notes for documenting and interpreting patient medical charts), as well as a series of four tests. The course also follows the recommended practice of tests that represented approximately half of the total course grade and varied assessments that include ongoing discussion, projects, case studies, and

applications. Quiz performance contributed to 40-50% of the overall grade across sections. The course had several sections taught completely online with different instructors administering the same curriculum. All six instructors agreed to use common exam formats that apply concepts from WCET's best practice for online education, including timed quizzes, random selection of questions from a common question pool, and responses that are in randomized order (WCET, 2009). In addition, students could not exit and restart an online quiz once they had begun, and could not view the exam after completing it.

All six instructors used the same proctoring software, Respondus Lockdown Browser™ + Respondus Monitor, a remote proctoring software that videotapes the student in their surroundings and also locks down their internet browser during the test so they cannot open other websites, nor can they take a screenshot, copy, or print exams to share them with others.

Participants and Responses Evaluated

Students enrolling in six sections of this class were analyzed. These students were from a variety of majors including the Kinesiology and Health (KNH) department, which were categorized under the following fields of study: Kinesiology (KNH-Kin), Health-Nutrition-Athletic Training (KNH-Health), Pre-Med, Business, Biological-Sciences (Bio-Sci) and others. For a full listing of majors within each category please see Appendix A.

Quiz scores and the time to complete each quiz from students were analyzed in this study. Each student completed four quizzes containing multiple choice questions pertaining to the four units in the course, with 60 minutes allowed to complete each quiz. The quizzes were administered through the online course management software, Canvas, and were uniquely generated for each student using questions randomly selected from question sets shared across all course sections. Questions were determined by the instructors to have similar difficulty levels.

Reasoning and Plan for Online Quizzes in Proctored and Non-Proctored Environments

A concern about integrity in distance learning is that due to the online administration of quizzes and tests, there is typically a heavy reliance on student honesty to refrain from using unauthorized reference materials during examinations. This especially applies in classes that have multiple sections, some of which may be offered online, while other sections are offered in a traditional format. The different ways of proctoring for online versus traditional tests may yield different results that do not accurately reflect student mastery of the content. This is particularly likely in closed-ended tests that measure recognition, understanding, and basic applications of information that can be easily looked up on the internet or in a textbook, rather than open-ended questions that involve more complex processing of information. There is a need in such cases to assure that the integrity of the course was upheld such that academic honesty of students was promoted to the best of the instructors' abilities.

Students were informed in writing about the following conditions and expectations that applied to all quizzes: students were to take these quizzes by themselves with no notes or other resources allowed during the exam. Students were not certain which of the quizzes would be proctored prior to the start of the exam. All quizzes covered similar material, and questions were randomly drawn from a shared question bank. Following the completion of proctored quizzes, thumbnails of the pre-quiz student photo, student ID, and

environment scan were generated, along with randomly timed thumbnails of the entire quiz video from Respondus Monitor. The thumbnails were available for review by the instructor of the course to detect rule violations or suspicious activity. The instructor could click on each thumbnail to view a brief timed interval of a portion of the video that recorded the student while taking the quiz in order to confirm whether or not a violation occurred during the exam.

Data Collection and Analyses

To explore the impact of proctoring software on student performance, the six course sections were assigned to a sequence of proctored/unproctored quiz progression. For shorthand reference to this sequence of proctored and unproctored exams, we will use a four character acronym of “P” and “U” in order of the quizzes (example: *UPUP* refers to a sequence with quizzes 1 and 3 being unproctored, and quizzes 2 and 4 being proctored) In designing the study, it was decided that the first two units contained easier materials than the last two units, so only the orders *PUUP*, *UPPU*, *UPUP*, and *PUPU* were considered to allow for one quiz of each proctor status in each half of the course.

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. No student requested removal of their quiz scores or other information. The anonymized data from all students who consented and had completed all four quizzes were then used in a statistical analysis to assess the effect of proctoring on exam scores and the percentage of allotted time taken.

Statistical Modeling of Quiz Scores

The effects of proctoring while controlling for the section, quiz, and major of the students were modeled. It was hypothesized that the four quiz scores from the same student will be naturally related, and also that the scores from students of the same section may be correlated as well. To accommodate this covariance structure, a mixed effects regression model (McLean, Sanders & Stroup, 1991) was used, with nested random effects for students and students within sections, and fixed effects to estimate the effect of proctoring status while controlling for student major and quiz number. The model promotes the most viable interpretability to a broader population of students as it acknowledges that the results are specific to the student majors and the quizzes in this particular Medical Terminology course.

The model selection process revealed that scores were significantly affected by proctoring, quiz number, and student major, and additionally that proctoring effects varied significantly across quizzes and majors; thus these were included in the fixed effects of the model. The nested random effects for students and students within sections were also found to provide stronger model fit. The selected mixed random effects model is therefore defined as:

$$\begin{aligned} \text{Score}_{jklmn} = & \text{Intercept} + \text{Proctor}_j + \text{Quiz}_k + \text{Major}_l + (\text{Proctor} \times \text{Quiz})_{jk} \\ & + (\text{Proctor} \times \text{Major})_{jl} + \gamma_m + \delta_{mn} + \epsilon_{jklmn} \end{aligned} \quad (1)$$

where Score_{jklmn} is the score for proctor status j on quiz k with major l from section m and student n . Random effect $\gamma_m \sim \text{Normal}(0, \sigma_m)$, $\delta_{mn} \sim \text{Normal}(0, \sigma_{mn})$, and $\epsilon_{jklmn} \sim \text{Normal}(0, \sigma)$ are assumed to be

independently distributed and represent the effect for class sections, effect for students within sections, and error terms for each individual quiz score, respectively.

Statistical Modeling of Time Taken

A similar modeling process to that described for quiz scores in the section above was used to explore the effect of proctoring on time taken for quiz completion, while controlling for other important factors. Model selection yielded a model with main effects for proctoring, quiz number, and student major, and interactions between quiz number and proctoring is included for time taken. Note that for time taken, the interaction between student major and proctoring status was not found to aid the model fit. The model fitting also suggests the need for nested random effects for students and students within sections. The final model fitted for time taken is defined as:

$$\text{Time}_{jklmn} = \text{Intercept} + \text{Proctor}_j + \text{Quiz}_k + \text{Major}_l + (\text{Proctor} \times \text{Quiz})_{jk} + \gamma_m + \delta_{mn} + \epsilon_{jklmn} \quad (2)$$

where Time_{jklmn} is the time taken for proctor status j on quiz k with major l from section m and student n . Random effect $\gamma_m \sim \text{Normal}(0, \sigma_m)$, $\delta_{mn} \sim \text{Normal}(0, \sigma_{mn})$, and $\epsilon_{jklmn} \sim \text{Normal}(0, \sigma)$ are assumed to be independently distributed and represent the effect for class sections, effect for students within sections, and error terms for each individual quiz time, respectively.

Results

Data Description

Proctoring is possible in the online setting if monitoring software is employed. In the plots in Figure 1 below, with proctor status (procStatus) indicated in red for unproctored and blue for proctored, we see that overall the proctored and unproctored quizzes start off with similarly high scores and short time taken on Quiz 1, but that a clear separation forms with unproctored quizzes scoring higher and taking longer on average. In Figure 2, with proctor status indicated as unproctored in red and proctored in blue, the bottom two graphs show that the overall trends of increasingly separated scores and time taken seem to hold for unproctored tests for most student majors; with the exception of the KNH-Health students outside of Kinesiology – including athletic training, nutrition, public health, and sports leadership and management. These KNH-health students had scores and times that were not easily separable by proctoring status. We can also note that all six of the Pre-Med students were in sections following the *PUUP* quiz progression and thus we will be careful with our interpretations for students of this group of majors. The top two plots in Figure 2 show that the unproctored quiz scores tended to be higher with longer time taken for most sections. Despite a miscommunication in administration of quizzes in a seventh section, the other sections (A-F) provide sufficient observations to allow the estimation of proctoring effects while controlling for quiz progression.

Interaction of Proctoring and Student Major on Online Test Performance
Alessio, Malay, Maurer, Bailer, and Rubin

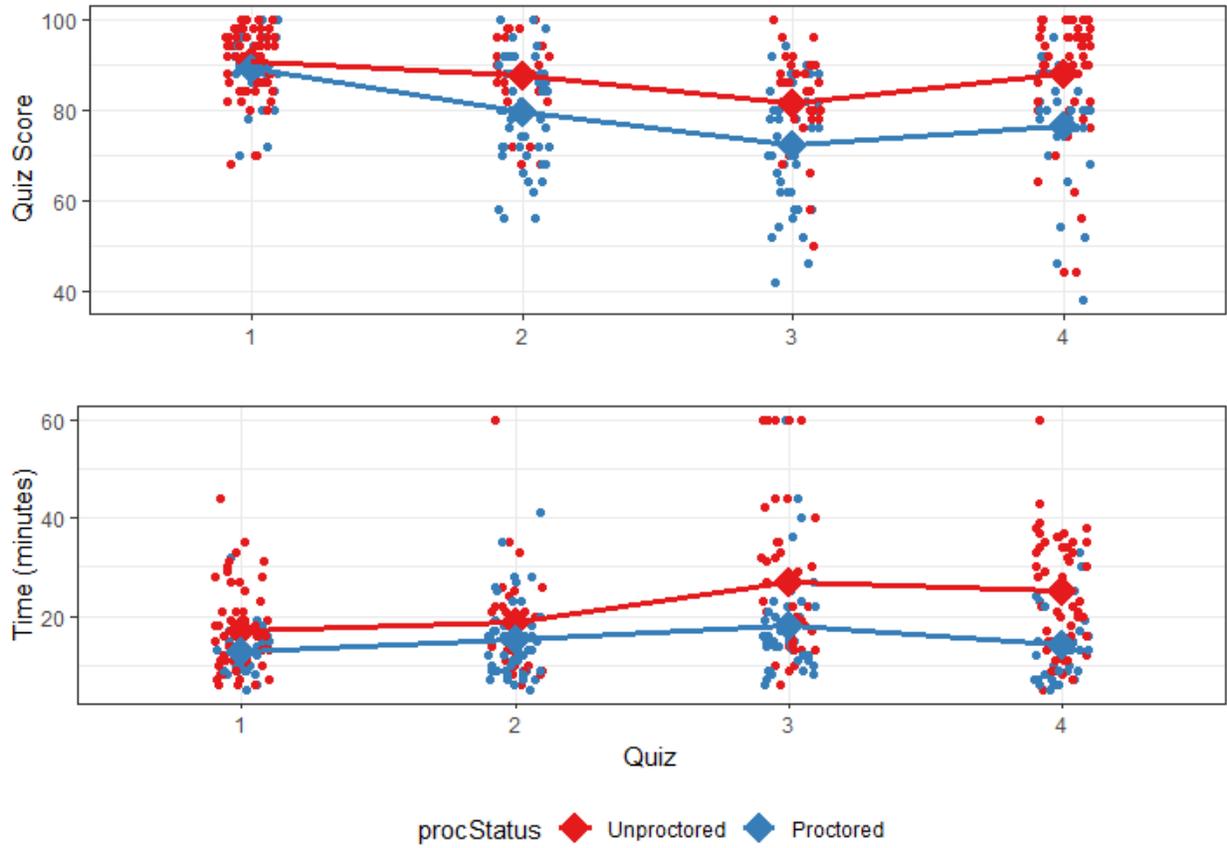


Figure 1. Quiz scores (top) and quiz times (bottom) – when unproctored and proctored.

Interaction of Proctoring and Student Major on Online Test Performance
Alessio, Malay, Maurer, Bailer, and Rubin

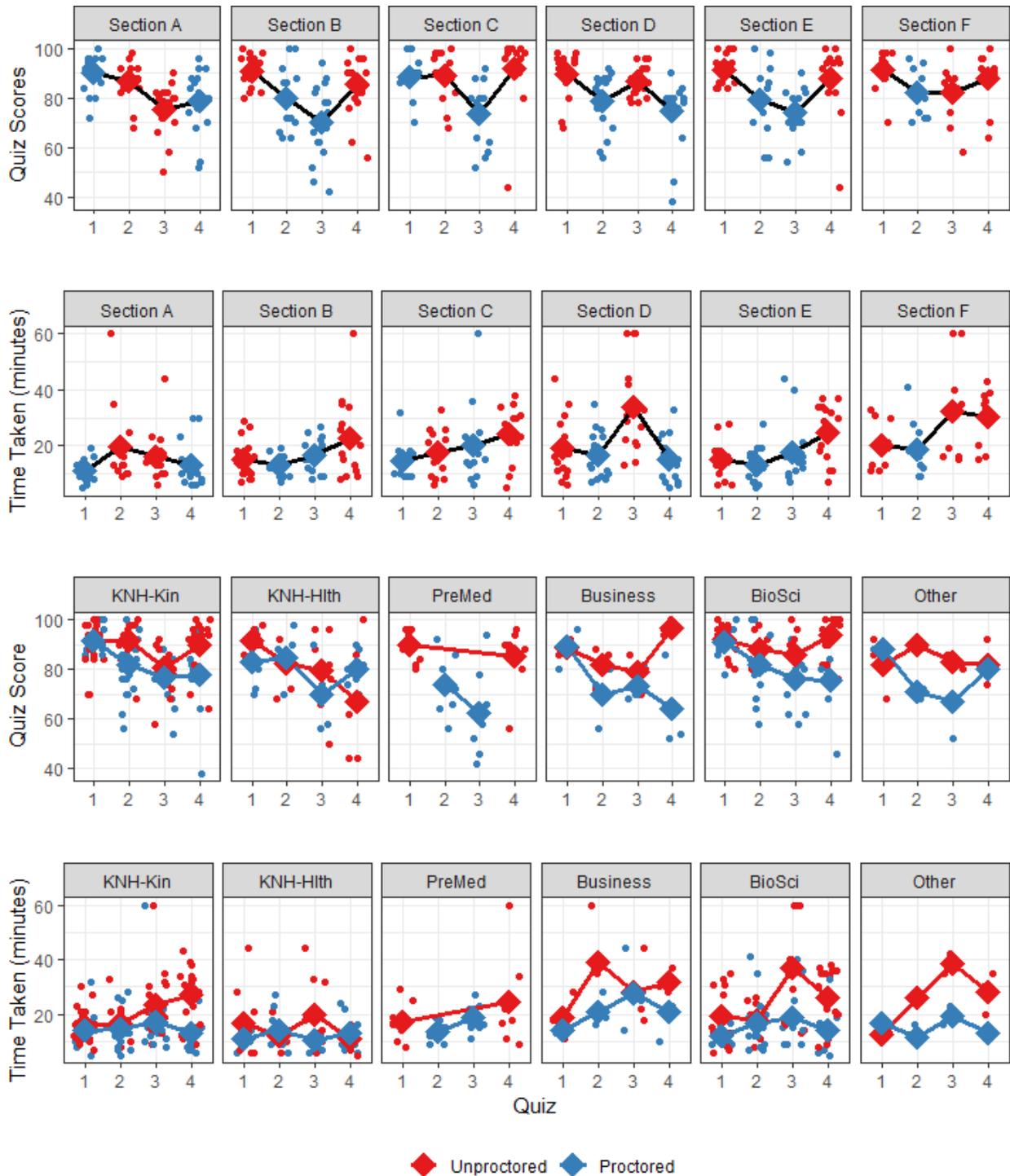


Figure 2. Four different quiz scores by section, quiz times taken by section A-F, quiz scores by major, and quiz times taken by major.

Modeling Quiz Scores

The fixed effects estimated from the fitted statistical model provide a few key insights about how proctoring, quiz progression, and student majors interact. Figure 3 displays the behaviors that are contained in the fixed effects structure of the fitted model for quiz scores. This plot shows predicted quiz scores for each student major and proctoring status with red lines representing unproctored and blue lines representing proctored status for each quiz. Model coefficients and associated inferential statistics are provided in Appendix B: please note that coefficients are interpreted relative to the intercept that represents a baseline of Kinesiology students (KNH-Kin major) unproctored on Quiz 1. We find that there is no statistically significant effect of proctoring for the baseline group on Quiz 1, however, through the interaction terms, we find that significant differences are manifested over the different quizzes and majors.

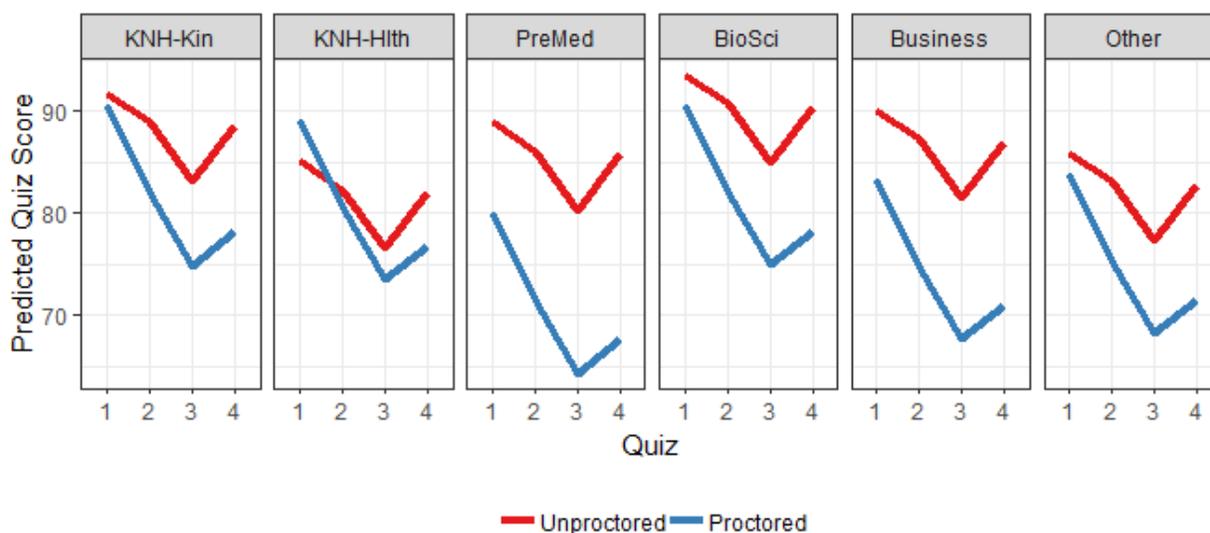


Figure 3. Plot of predicted quiz scores under each combination of student major, quiz number, and proctoring status based on fixed effects structure from fitted model.

Not all quizzes are equally difficult, with statistically significantly different average scores even while controlling for majors/proctoring. Quiz 3 appears to be the most difficult, with an estimated 15 point lower score than on Quiz 1. It is also clear that in nearly all cases, the proctored scores are predicted to be lower than unproctored scores. On Quiz 1 and 2 the estimated scores for unproctored are not statistically significantly higher, but this difference grows to a statistically significant 7 point and 9 point higher score for unproctored students on Quizzes 3 and 4, respectively.

The interaction terms show some distinctly different results of proctoring within the student majors as well. Majors in Bio-Sci, Business, and Other displayed no statistically significant differences from the baseline KNH-Kinesiology cohort, each having notably lower scores when proctored on Quizzes 3 and 4. However, the KNH-Health showed significantly lower unproctored scores than the baseline, while having a positive adjustment when being proctored; the net effect overall is that they are the only cohort without a major change in scores when being proctored. Also apparent is the gap between proctored and unproctored quizzes, which is statistically significantly larger for pre-med students than the rest of the majors.

Modeling Time Taken

We now take a similar approach for interpreting the proctoring effects on the time taken per quiz, based on the model discussed in the Methods section. In Figure 4 below, we plot the predicted time for quiz completion under each major, with proctoring indicated by red lines for unproctored and blue for proctored, and quiz combination that represent the fixed effects portion of the model for time taken. The model of coefficients and associated inferential statistics relevant to this figure is provided in Appendix C.

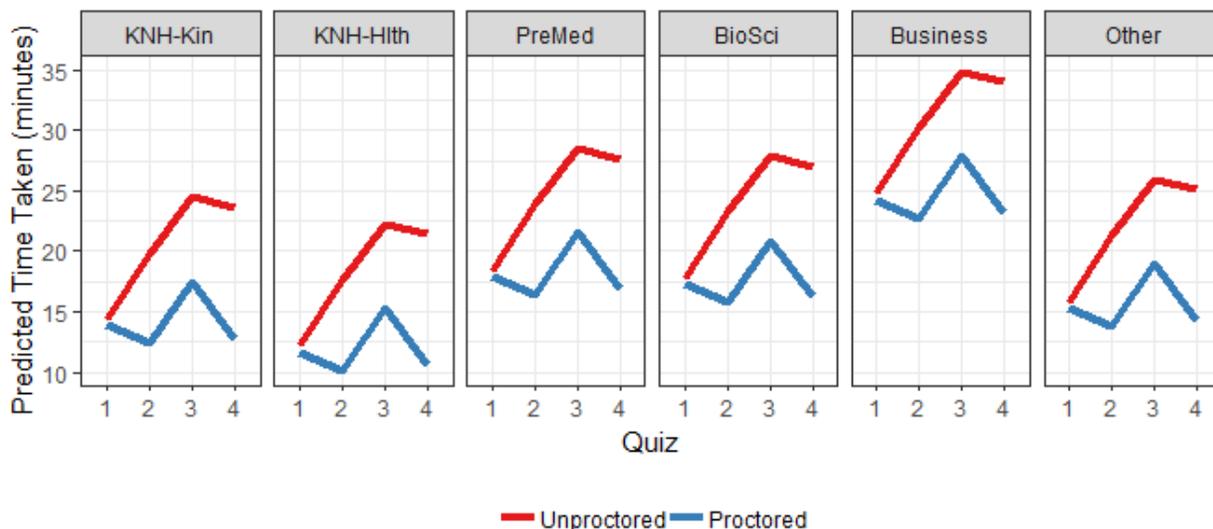


Figure 4. Plot of predicted time to complete test under each combination of student major, quiz number, and proctoring status based on fixed effects structure from fitted model.

From the model, it is shown that in Quiz 1 there is no statistically significant effect for proctoring on time taken; however, on all subsequent quizzes there was a statistically significant increase in the time taken when the quizzes were unproctored. It is additionally noted that only the business students had statistically significant difference in time taken from the baseline of KNH-kinesiology students, taking an estimated 10.4 minutes longer per quiz.

Discussion

This study was conceived and conducted as a structured randomized design that reported significantly different online quiz scores, as well as time taken to complete online quizzes when comparing students who were tested with and without proctoring software and audiovisual monitoring. The number of quizzes, length of quiz time, and number of quiz questions that were proctored and unproctored were similar for each section. Based on Beck's (2014) suggestion that multiple factors may influence academic dishonesty, including student major, this study compared test results and time taken among a variety of student majors. The first research question inquired about the effects of proctoring software on students' test scores and quiz completion time in comparison to quizzes without proctoring. Results showed that the unproctored quiz scores tended to be higher with longer time taken for most sections compared with proctored quiz scores. Since these results happened both across and within sections, together, these two findings suggest

that when not proctored, students may spend extra time looking up answers using resources that were not allowed during the test.

In our comparison of proctoring effects by student major, a few groups were identified to have significantly different behavior than the rest. Significant differences in grade disparity were found with students who were pre-med and business majors when comparing proctored vs. unproctored and time taken to complete the unproctored tests, respectively. It is noteworthy that the students in these major categories may feel higher academic pressures for high grade than peers in other majors; from medical schools admissions requiring a high GPA and the business college requiring a minimum GPA of 3.4/4.0 for acceptance. Based on results of this study, which controlled for exam difficulty, course design, instructor effects, and student majors, the main findings are that online quiz scores were significantly affected by a) proctoring, b) student major, and c) quiz progression or difficulty.

In addressing the effect of exam difficulty on scores in online tests, with and without the use of proctoring software, results showed that mean grades for the first quiz, across all sections and majors, were highest compared with all other quizzes, indicating a difference in level of difficulty. Grade discrepancy between proctored and unproctored for quiz 1 was the smallest, except for those who identified as “pre-med” students. The grade discrepancies increased with the level of difficulty in each subsequent quiz. The third and fourth quizzes appeared to be the most difficult, and also showed the largest grade disparities in all majors, especially those in pre-med. The overall proctoring effect on the more difficult quizzes approximated 7-9 points lower when students were not proctored compared with when they were proctored.

Time discrepancy between proctored and unproctored quizzes was greater for business students who used much more time to complete their tests compared to all other students. These were the two groups that arguably had the highest stakes for earning a high grade: admission into medical school and admission to, or continuation in, business school. This supports the triangle model of cheating used by D’Souza and Siegfeldt (2017), indicating that factors that increase the incentive to cheat, the opportunity to cheat, and provide rationalization for cheating, all will lead to dishonest behavior.

These findings contradict those of Beck (2014), who found that online unproctored exams were not different from face-to-face, proctored exams. However, Beck’s study identified several limitations including a very small sample and possible instructor effects from a single professor. This study controlled for these problems, as well as others such as grade inflation, changes over time (year to year), a test design that does not reflect best practices for discouraging cheating, instructor behavior, and exam difficulty. The findings of the current study support those of other studies (e.g., Alessio et al., 2017; Kitahara et al., 2011) who recommended webcam proctoring in all online courses. The content of the course investigated in this study, Medical Terminology, aligns closely with content expected in nursing courses examined by Mirza and Staples (2010). Therefore it was not surprising that the current results of disparate grades on proctored and unproctored tests, and implications support those of Mirza and Staples (2010), who further reported that students said they were less likely to cheat when monitored with the webcam during online testing.

Conclusions

Academic dishonesty has occurred long before online learning and testing were introduced to academia. The difference in addressing cheating in online classes includes unique challenges that new technology presents that differ from face-to-face situations (Christie, 2003). Instructions and expectations for academic honesty are often written in similar language for both traditional and online courses. However, when students are separated from their instructors and do not experience personal communication, including tone, sense of presence, and facial expressions, they view cheating differently and less negatively than in traditional settings (Moten, Fitterer, Brazier, Leonard, & Brown, 2013). In an effort to get an advantage, some students in online courses turn to dubious businesses that sell academic papers, develop software that assists in cheating during online tests, and even arrange for third-party test takers. Students perceive that their risk of getting caught is low, consequences are light, and have reasons that include a desire to help others as well as themselves (Christie, 2003). Self-reporting of cheating is difficult to interpret due to different survey results ranging from online students reporting they cheated less than face-to-face students (Kidwell & Kent, 2008) to students who admit they were more than four times as likely to cheat when taking an online vs. face-to-face class (Watson & Sottile, 2010).

In a study that explored academic dishonesty beyond self-report data, Alessio et al. (2017) conducted a natural study design that compared online test results from proctored versus unproctored online tests in nine sections of the same course. When proctoring software that included audiovisual monitoring was used, the average test grade was 17 points lower compared with students who were not monitored. This grade disparity occurred both within the same class, when students were proctored on one test and not proctored on another. It also occurred between classes, comparing students who were proctored in one section with students who were not proctored in another section Alessio et al., 2017]; Kitahara et al., 2017).

The current study was conceived and conducted as a structured randomized design that provided a higher level of confidence in the findings. It found significantly different online quiz scores, as well as time taken to complete online quizzes when comparing students who were tested with and without proctoring software and audiovisual monitoring. Grade disparities were observed in the most difficult quizzes and were particularly large for students who identified as pre-medical studies. Compared with all other majors, students who identified as majoring in business used the most time to take a quiz when proctored. Overall, use of proctoring software resulted in lower quiz scores, shorter quiz taking times, and less variation in quiz performance across exams, implying greater compliance with academic integrity compared with when the quiz was taken without proctoring software. These results affirm the value of using proctoring software for online tests and quizzes, especially when exam difficulty progresses over time, and to address the uneven performances by student major.

Limitations and Future Study

This study was limited to a particular course that has a high degree of memorization of terms. While there were multiple ways of assessing student learning in this course (e.g., discussions, case studies), tests primarily included objective questions, with one best answer. The results may not be generalized to more

broad-based courses that incorporate theory, calculations, and subjective type questions. Future studies should examine a wide range of courses in a variety of majors that reflect a wider breadth of assessment.

Cheating, whether it is planned or acutely panic driven, results in students violating test taking rules (Bunn, Caudill, & Gropper, 1992), which leaves the burden on faculty and administrators to prevent, detect, and when appropriate, hold accountable, students who engage in academic dishonesty. This is no small task, with estimates of undergraduate cheating that ranges from 30% to 96% of students (Nonis & Swift, 2001). Identifying when cheating occurs is time consuming and stressful as it requires instructors and proctors to gather evidence of the infraction. An allegation of academic dishonesty then follows an established protocol at the institution that includes due process with evidence presented and all sides heard. Ultimately, a decision is made on whether or not academic dishonesty occurred followed by an appropriate disciplinary action, such as exoneration, or warning, suspension, or expulsion of students found responsible of academic dishonesty. Universities would benefit from systematic integrity practices that include clear preventative guidelines to faculty and students, as well as products designed to prevent academic dishonesty, so that academic integrity can be assured using the best evidence-based strategies.

References

- Alessio, H.M., Malay, N.J., Maurer, K.T., Bailer, A.J., & Rubin, B. (2017). Examining the effect of proctoring on online test scores. *Online Learning*, 21(1), 146-161. <http://dx.doi.org/10.24059/olj.v21i1.885>.
- Amigud, A., Arnedo-Moreno, J., Daradoumis, T., & Guerrero-Roldan, A. (2017). Using learning analytics for preserving academic integrity. *The International Review of Research in Open and Distributed Learning*, 18(5). <http://dx.doi.org/10.19173/irrodl.v18i5.3103>
- Beck, V. (2014). Testing a model to predict online cheating - much ado about nothing. *Active Learning in Higher Education*, 15(1), 65-75. <https://doi.org/10.1177/1469787413514646>
- Berkey, D., & Halfond, J. (2015). Cheating, student authentication and proctoring in online programs. *New England Board of Higher Education, Summer*, 1-5. Retrieved from <http://www.nebhe.org/thejournal/cheating-student-authentication-and-proctoring-in-online-programs/>
- Boehm, P.J., Justice, M., & Weeks, S. (2009). Promoting academic integrity in higher education [Electronic document]. *The Community College Enterprise*, Spring, 45-61. Retrieved from <https://pdfs.semanticscholar.org/bd69/cc9738a66f10d1e049261659faba1efedc87.pdf>
- Bunn, D.N., Caudill, S.B., & Gropper, D.M. (1992). Crime in the classroom: An economic analyses of undergraduate student cheating behavior. *The Journal of Economic Education*, 23(3), 197-207. doi:10.1080/00220485.1992.10844753
- Christie, B. (2003). Designing online courses to discourage dishonesty: Incorporate a multilayered approach to promote honest student learning, *Educause Quarterly*, 26(4), 54-58. Retrieved from <https://er.educause.edu/~media/files/articles/2003/10/eqm0348.pdf?la=en>
- Cluskey, G.R., Ehlen, C.R., & Raiborn, M.H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics*, 4, 1-7. Retrieved from <http://www.aabri.com/manuscripts/11775.pdf>
- 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. <http://dx.doi.org/10.1145/2810239>
- D'Souza, K.A., & Siegfeldt, D.V. (2017). A conceptual framework for detecting cheating in online and take-home exams. *Decision Sciences Journal of Innovative Education*, 15(4), 370-391. <https://doi.org/10.1111/dsji.12140>
- Etter, S., Cramer, J.J., & Finn, S. (2006). 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. Retrieved from

<http://www.eacfaculty.org/pchidester/Eng%20102f/Plagiarism/Origins%20of%20Academic%20Dis%20honesty.pdf>

- Fask, A., Englander, F., & Wang, Z. (2014). Do online exams facilitate cheating? An experiment designed to separate out possible cheating from the effect of the online test taking environment. *Journal of Academic Ethics, 12*(2), 101-112. doi:10.1007/s10805-014-9207-1
- Kidwell, L.A., & Kent, J. (2008). Integrity at a distance: A study of academic misconduct among University students on and off campus. *Accounting Education, 17*(1), S3-S16.
<https://doi.org/10.1016/j.sbspro.2014.01.1141>
- 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. Retrieved from <https://eric.ed.gov/?id=EJ904058>
- Kitahara, R., Westfall, F., & Mankelwicz, J. (2011). New, multifaceted hybrid approaches to ensuring academic integrity. *Journal of Academic and Business Ethics, 3*, 1-12. Retrieved from <http://www.aabri.com/manuscripts/10480.pdf>
- McCabe, D.L., Trevino, L.K., & Butterfield, K.D. (2001). Cheating in academic institutions: A decade of research. *Ethics and Behavior, 11*(3), 219-232. doi:10.1207/S15327019EB1103_2
- McGee, P. (2013). Supporting academic honesty in online courses. *Journal of Educators Online, 10*(1). Retrieved from https://www.thejeo.com/archive/2013_10_1
- McLean, R.A., Sanders, W.L., & Stroup, W.W. (1991). A unified approach to mixed linear models. *The American Statistician, 45*(1), 54-64. doi:10.1080/00031305.1991.10475767
- Mirza, N. & Staples, E. (2010). Webcam as a new invigilation method: Students' comfort and potential for cheating. *Journal of Nursing Education, 49*(20), 116-119. doi:10.3928/01484834-20090916-06
- Moten, J. Jr, Fitterer, A., Brazier, E., Leonard, J., & Brown, A. (2013). Examining online college cyber cheating methods and prevention methods. *The Electronic Journal of e-Learning, 11*(2), 139-146. Retrieved from <https://eric.ed.gov/?id=EJ1012879>
- Nonis, S., & Swift, C. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty. A multi-campus investigation. *Journal of Education for Business, 77*(2), 69-77.
<https://doi.org/10.1080/08832320109599052>
- Northcutt, C.G., Ho, A.D., & Chuang, I.L. (2016). Detecting and preventing "multiple-account" cheating in massive open online courses. *Computers and Education, 100*, 71-80. Retrieved from <http://arxiv.org/abs/1508.05699>
- Rogers, C.F. (2006). Faculty perceptions about e-cheating during online testing. *Journal of Computing Sciences in Colleges, 22*(2), 206-212.

- Rubin, B., & Fernandes, R. (2013). Measuring the community in online classes. *Online Learning Journal*, 17(3), 115-135.
- Rujoiu, O. & Rujoiu, V. (2014). Academic dishonesty and workplace dishonesty. An overview. *Proceedings of the 8th International Management Conference, Faculty of Management, Academy of Economic Studies, Bucharest, Romania*, 8(1), 928-938. Retrieved from <https://ideas.repec.org/a/rom/mancon/v8y2014i1p928-938.html>
- Schuhmann, P.W., Burrus, R.T., Barber, P.D., Graham, J.E., & Elikai, M.F. (2013). Using the scenario method to analyze cheating behaviors. *Journal of Academic Ethics*, 11, 17-33.
<https://doi.org/10.1080/10508422.2015.1051661>
- 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), 1-13. Retrieved from <https://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-AcademicIntegrity-2009.pdf>

Appendix A

List of Student Majors in Each Major Category

Field of study	Included majors	Number of students
KNH-kin	Kinesiology	40
BioSci	Biology, Zoology, Microbiology, Ecology, Chemistry, Biochemistry, and Psychology	22
KNH-health	Sports Leadership and Management, Public Health, Nutrition, Athletic Training, and Healthcare Professionals	17
PreMed	Biochemistry, Microbiology, Biology, University Studies or Healthcare Professionals <i>with a declared pre-med emphasis.</i>	8
Business	Accounting, Marketing, Supply Chain and Operations Management, and Finance	6
Other	Speech Pathology and Audiology, Interactive Media Studies, and University Studies	4

Appendix B

Fitted Model Coefficients for Student Quiz Scores Model

Model term	Value	SE	DF	t-value	p-value	
(Intercept)	91.78	1.936	240	47.416	< 0.001	
Proctored	-1.184	2.618	240	-0.452	0.651	
Quiz 2	-2.831	2.283	240	-1.240	0.216	
Quiz 3	-8.601	1.906	240	-4.512	< 0.001	*
Quiz 4	-3.101	1.702	240	-1.822	0.070	*
Proctored* Quiz 2	-5.742	3.557	240	-1.614	0.108	
Proctored* Quiz 3	-7.136	3.169	240	-2.252	0.025	*
Proctored* Quiz 4	-9.276	3.031	240	-3.061	0.002	*
KNH-Hlth	-6.641	2.936	73	-2.262	0.027	*
PreMed	-2.854	3.664	73	-0.779	0.438	
Business	-1.647	4.062	73	-0.405	0.686	
BioSci	1.850	2.519	73	0.735	0.465	
Other	-5.815	4.742	73	-1.226	0.224	
Proctored* KNH-Hlth	5.283	2.819	240	1.874	0.062	**
Proctored* PreMed	-7.711	3.534	240	-2.182	0.030	*
Proctored* Business	-5.611	3.868	240	-1.450	0.148	
Proctored* BioSci	-1.761	2.456	240	-0.717	0.474	
Proctored* Other	-0.892	4.695	240	-0.190	0.849	

Note. *denotes a statistically significant result at the $\alpha = 0.05$ level; **denotes a statistically significant result at the $\alpha = 0.1$ level.

Appendix C

Fitted Model Coefficients for Student Quiz Time Taken Model

Model term	Value	SE	DF	t-value	p-value	
(Intercept)	14.354	2.286	245	6.279	< 0.001	
Proctored	-0.491	1.937	245	-0.254	0.800	
Quiz 2	5.473	1.937	245	2.826	0.005	*
Quiz 3	10.106	1.564	245	6.461	< 0.001	*
Quiz 4	9.267	1.383	245	6.703	< 0.001	*
Proctored* Quiz2	-6.999	3.151	245	-2.221	0.027	*
Proctored* Quiz 3	-6.453	2.708	245	-2.383	0.018	*
Proctored* Quiz 4	-10.353	2.522	245	-4.104	< 0.001	*
KNH-Hlth	-2.177	2.565	73	-0.849	0.399	
PreMed	4.020	3.584	73	1.122	0.266	
Business	10.396	3.571	73	2.911	0.005	*
BioSci	3.391	2.191	73	1.548	0.126	
Other	1.482	4.067	73	0.365	0.717	

Note. *denotes a statistically significant result at the $\alpha = 0.05$ level; **denotes a statistically significant result at the $\alpha = 0.1$ level.

