

Cheating on Unproctored Online Exams: Prevalence, Mitigation Measures, and Effects on Exam Performance

Jacob Pleasants
University of Oklahoma, USA

John M. Pleasants
Barbara P. Pleasants
Iowa State University, USA

Abstract

As online courses become increasingly common at the college level, an ongoing concern is how to ensure academic integrity in the online environment. One area that has received particular attention is that of preventing cheating during unproctored online exams. In this study, we examine students' behavior during unproctored exams taken in an online introductory biology course. A feature of the learning management platform used for the course gave us the ability to detect cheating behavior involving students leaving the test page and viewing other material on their computers. This allowed us to determine what proportion of students cheated and examine the efficacy of various measures to mitigate cheating. We also explored the relationship between cheating behavior and exam performance. We found that 70% of students were observed cheating, and most of those who cheated did so on the majority of test questions. Appealing to students' honesty or requiring them to pledge their honesty were found to be ineffective at curbing cheating. However, when students received a warning that we had technology that could detect cheating, coupled with threats of harsh penalties, cheating behavior dropped to 15% of students. Unexpectedly, we did not find evidence that students' exam performance changed when their cheating behavior changed, indicating that this common form of cheating might not be as effective as students, or their instructors believe it to be.

Keywords: cheating, online exams, undergraduate, proctor, academic honesty

Pleasants, J., Pleasants, J. M., & Pleasants B. P. (2021). Cheating on unproctored online exams: Prevalence, mitigation measures, and effects on exam performance. *Online Learning*, 26(1), 268-284. DOI: 10.24059/olj.v26i1.2620

Ensuring post-secondary students' academic integrity is a long-standing concern of colleges and universities. With the proliferation of online resources and online coursework, maintaining high standards for academic honesty has become increasingly complex (Spaulding, 2009). The information and communication technologies that have enabled online education are boons in many respects, but they have also given students new and powerful means to engage in dishonest behavior (Dyer et al., 2020; Stogner et al., 2013; Watson & Sottile, 2010).

Student behavior during online unproctored testing is an area of particular concern for college faculty (McNabb & Olmstead, 2009)—a concern made even more relevant by the COVID-19 pandemic when unproctored online tests became the norm. In a recent survey conducted by Wiley (2020), 93% of instructors indicated a belief that students were more likely to cheat on online unproctored tests than on proctored tests. Those concerns are supported by a growing body of empirical work that has found evidence of cheating during unproctored online exams (Alessio et al., 2017, 2018; Fask et al., 2014; Hylton et al., 2016). Many instructors and institutions are therefore turning toward technologies such as Lockdown Browser (Respondus, 2020a) or webcam-based monitoring services that enable remote proctoring (e.g., Respondus, 2020b). However, those technologies come with substantial drawbacks in that they are both costly and invasive (Flaherty, 2020).

The present study examines the problem of cheating during unproctored online exams in the context of an undergraduate introductory biology course. We investigate the prevalence of cheating on the exams in the course and the extent to which various non-invasive measures inhibited cheating. A unique aspect of this study is that we were able to detect whether, during an exam, students navigated away from the test webpage and viewed other pages or documents open on the desktop. Viewing unauthorized materials is a particularly common form of cheating (Stephens et al., 2007), and we were able to determine how different mitigation strategies affected the prevalence of that cheating behavior. We also explore how cheating behaviors are associated with test performance. We address the following research questions:

- 1) What percentage of students exhibit cheating behaviors when taking tests in an unproctored environment?
- 2) What percentage of students exhibit cheating behaviors when (a) an appeal is made to their conscience to uphold academic integrity, (b) they have to sign an honesty pledge, or (c) are told they are being surveilled?
- 3) How are cheating behaviors related to test performance?

Literature Review

Studying students' cheating behavior during tests is inherently difficult. Direct observational evidence for student cheating is often difficult to obtain, and students have good reason not to admit to cheating, even on anonymous surveys (Kervliet & Sigmann, 1999). Surveys of students in online courses have not always indicated that online environments lead to more cheating than face-to-face ones, although there is some evidence that students are more likely to consult unauthorized materials during online exams (Grijalva et al., 2006; Stephens et al., 2007; Stuber-McEwen et al., 2009; Watson & Sottile, 2010). A recent survey by Dyer, Pettyjohn, and Saladin (2020) highlights that concern and also raises the importance of proctors during exams. They examined student reports of cheating behavior in proctored and unproctored settings, as well as students' beliefs about the acceptability of various dishonest behaviors. Notably, they found that students viewed certain dishonest behaviors, including looking up

answers in unauthorized materials, as *more acceptable* in unproctored settings. Many students seemed to believe that a lack of a proctor meant that the instructor was not serious about certain resources being “off limits.”

Although surveys of students can be informative, investigations of cheating that go beyond self-reports are essential. In an unproctored online setting, directly observing student cheating behavior is naturally quite challenging, barring the use of surreptitious monitoring (Kervliet & Sigmann, 1999). Researchers who have investigated this phenomenon have therefore typically used student exam scores as an indicator of possible cheating behavior. If exam scores for students taking online unproctored exams are higher than those for students in proctored settings, then cheating is inferred.

Hollister and Berenson (2009) compared exam performance between two sections of the same course in which the only difference was that students in one section took exams in-person with proctors whereas the other took the exams online without a proctor. After controlling for a variety of covariates, they found no differences between the performances of the two sections. Beck (2014) similarly found that while variables such as students' GPA was predictive of test scores, the presence of proctors was not. However, a carefully controlled study by Fask, Englander, and Wang (2014) reached the opposite conclusion. They reasoned that in order to compare an in-person proctored exam with an online unproctored exam, the test setting (classroom versus home) also needs to be considered in order to discern the proctor effect. After controlling for setting, Fask, Englander, and Wang found evidence of elevated scores among students in the unproctored group, which they attributed to cheating behavior.

In recent years, technologies have been developed that enable online exams to be proctored even when taken from home. Typically, these technologies involve using webcams and/or screen-sharing to monitor student behavior during an online exam (Dunn et al., 2010; Flaherty, 2020; Grajek, 2020). Recent studies have investigated the impact of those technologies on student exam performance in online courses. Hylton, Levy, and Dringus (2016) randomly assigned students in an online course to an unproctored or webcam-based proctoring condition during exams. They found that students in the unproctored group had elevated exam scores and also took longer to complete their exams. The same findings were obtained in a sequence of studies by Alessio et al. (2017, 2018), who also studied the effects of webcam-based proctoring on the exam performance of online students.

The above studies suggest that webcam-based proctoring technologies are effective in reducing cheating behavior, but there remain multiple unresolved issues. Hylton, Levy, and Dringus (2016) as well as Alessio et al. (2017, 2018) found that students took longer to complete unproctored exams, but the extent to which that finding is indicative of cheating is not clear. Hylton, Levy, and Dringus (2016) point out the ambiguous role of test time and argue for its further study. This is particularly important because tightly limiting students' time to complete online exams is often suggested as a method of curtailing cheating (e.g., Cluskey et al., 2011). In addition, even if webcam-based proctoring technologies inhibit cheating, they are costly for institutions to implement and are disliked by students due to their invasiveness (Flaherty, 2020; Grajek, 2020). That invasiveness itself might reduce student test performance by making students nervous and uncomfortable (Hylton et al., 2016). Finally, the research on student behavior during online exams does not indicate how widespread cheating behavior is. Although studies have found elevated test performance *on average* during unproctored exams, what is unclear is what *proportion* of students are driving that elevation.

Overall, if a goal is to curtail student cheating during online exams, webcam-based proctoring is potentially effective but heavy-handed. Given the proliferation of online courses, the phenomenon of cheating needs to be better understood before costly technologies are deployed. At the same time, it is worth investigating whether less costly and less invasive options might also be effective in curtailing cheating behavior. As noted above, one common suggestion is to limit the amount of time students have to complete online exams (Cluskey et al., 2011; McGee, 2013). Another low-cost option is to have students pledge their adherence to academic honesty at the beginning of each online exam. Prior studies suggest that honesty is promoted by requiring participants to make affirmations of their honesty prior to engaging in tasks where cheating is likely to occur (e.g., Mazar et al., 2008).

Contribution of the Present Study

Many of the studies reviewed above rely on the assumption that elevated test scores (and in some cases, test times) are indicative of cheating. On its face, that is a reasonable assumption, but it treats student behavior in aggregate and as a black box, one that we aim to open up in the present study. In this study, we examine the test-taking behavior and performance of students in an undergraduate online biology course who completed exams without a proctor. We were able to detect the test-taking behavior of individual students using an Action Log created by the learning management software used in the course: Canvas (Instructure, 2020). The most likely way for students to cheat in an unproctored setting is to search the internet or view electronic notes on their computer. The Action Log provides data on when a student leaves the test page and examines other material.

We use the Action Log data to illuminate several important issues. First, we examine the prevalence of dishonest student behavior after several different non-invasive measures were implemented to attempt to curtail it. These measures were non-invasive in that they did not involve webcam-based monitoring of student behavior, nor the installation of any specific software. Second, we examine how students' engagement in cheating behaviors was related to their test performance. Because we are able to examine students' behavior at the individual level, we can more effectively investigate that relationship by not relying on aggregate performance.

Methods

Context

This study examines an online undergraduate introductory biology course at a large research university located in the Midwestern United States. The study was motivated by the university's response to the COVID-19 pandemic in the spring of 2020. Midway through that semester, students were sent home to complete their courses. At the beginning of that semester, students enrolled in the online biology course took their exams in the university testing center with a proctor present. After the students went home, all exams were taken on their own, without a proctor. We were naturally concerned about the possibility of cheating during those unproctored exams, and we noticed a marked increase in students' test scores after they were sent home. To more carefully investigate that phenomenon, we designed the present study to take place during the online course that ran during summer of 2020.

The biology course has been taught by the second author completely online for many years. It is an introductory-level course required for many science majors and the first of a two-course sequence. In summer 2020, 66 students completed the course, 23% of whom were freshmen, 37% sophomores, 28% juniors and 12% seniors. The course is taught completely asynchronously. The lecture materials in the course consist of presentation slides with voice-over

narration. The text portion of the slides is compiled into lecture notes that are electronically provided to the students along with portable document files of all the presentation slides.

The course has 8 exams, all of which are delivered within the Canvas Learning Management System (Instructure, 2020). The summer course runs for 12 weeks and there are 4 testing deadlines, occurring every 3 weeks. The first 2 exams must be done by the first deadline, the second two exams by the second deadline, and so on. Each pair of tests remains open for the entire 3 week period. Each exam has 20 questions drawn from a bank of over 100 questions and includes a mixture of multiple-choice and short-answer questions. Short-answer questions require students to input a few words or sentences in a text box. The multiple-choice questions are machine graded; the short-answer questions are graded by a teaching assistant. Although the mix of multiple-choice and short-answer questions varies by exam, on average less than 10% the questions are short-answer.

Exam Conditions

Under normal circumstances, the exams in the course are taken at a university testing center with a proctor present or, if the student is not on campus, with an approved proctor present. During the summer of 2020, students took all of their exams from home without a proctor. Given our concerns about potential cheating, we decided to try several measures to limit cheating behavior.

For the first exam, we split the course into two equal-sized groups using random assignment. One group (the “Appeal” group) was sent the following message at the beginning of the course, and the message was included as a header on the first exams:

It is important for the integrity of this course, the meaningfulness of grades, and fairness to other students that you do not use notes or any other materials while taking these tests.

The other group (the “Pledge” group) was required to respond true/false to a statement at the beginning of the first exam. The statement was: “I have not used notes or any other material while taking this test.” For Exams 2 to 4, all students were assigned to the “Pledge” condition.

Second, to see if more restrictive time limits on tests could curb cheating, we imposed tight time constraints on the first two exams for all students. For Exams 1 and 2 in summer 2020, we set the time equal to the historical mean for proctored tests plus one standard deviation, for a time limit of 20 minutes. For Exams 3 and 4, we loosened the time restriction; the time limit was set equal to the historical mean plus 2.5 standard deviations, or 30 minutes for Exam 3 and 40 minutes for Exam 4.

Midway through the semester, after Exams 1 to 4 were completed, we found that none of our measures were effective at curbing cheating behaviors. We therefore instituted a third approach for the remaining four tests: a stronger warning coupled with a notification of surveillance. All students were sent the following message:

This is a warning that due to concerns about students cheating on tests we now have the capability of monitoring student activity while taking tests. If I detect suspicious behavior on any of the remaining tests, I will have to take administrative action.

REPLY TO THIS E-MAIL TO LET ME KNOW YOU UNDERSTAND THIS WARNING.

That message was then placed as a header on every test (minus the third statement requesting a reply via email). Importantly, the statement was deliberately vague about how students were being monitored. Students may have thought that they were being observed via their webcam or some other unknown means. We reasoned that if students knew exactly how they were being monitored (and how they were not), they might simply cheat in ways that they knew we could not detect. By using vague language, we hoped to reduce cheating in general rather than just one specific means of cheating. In addition, for students whose Action Logs still showed cheating behavior on Tests 5 and 6, the instructor sent the following email message:

[Student name], I noticed that you have had other web pages open when you are taking exams. You must have just the test webpage open and remain on that page while you are taking an exam. If you are accessing notes on other pages during the test, I can't be certain of the tests' validity. If I see evidence of this on the remaining exams, I will be forced to give you zeros.

Data Collection

The data collected for the present study include students' scores on the eight exams, times to complete each of the exams, and Action Logs of students' behavior on the exams, described more extensively below. All data were anonymized by the instructor before analysis.

Characterizing Student Behavior

When an online exam is completed within Canvas, an Action Log is created that records a student's activity during an exam. It creates a time stamp when a student answers a question as well as when a student leaves the test page to view another page. A detailed guide describing the data produced by the Action Logs and how we interpreted them is included in the supplemental materials. The Action Logs provide an indication of cheating because the most likely way for a student to cheat is to consult disallowed materials on their computer (such as a website or the lecture notes that they were provided). Doing so, however, would require that the student navigate away from the exam page, which would be recorded in the Action Log. Of course, not all cases of leaving the test page are necessarily instances of cheating; a student might, for instance, be answering an email or responding to a social media message. Repeated instances of leaving the test page, however, are unlikely to be so benign.

Operationally, we defined an instance of "cheating on an exam question" as occurring when the Action Log indicated that a student had left the exam page prior to answering that question. If there were no instances of leaving the test page between a student answering a question and having answered the previous one, then we defined that as a non-instance of cheating. The vast majority of exam questions were multiple-choice, but some tests had one or more short-answer questions that required students to type a few words or sentences into a text box on the test page. We excluded short-answer questions from analysis because certain web browsers create false instances of leaving the test page when students type into a text box.

For each exam taken by each student, we determined the "Extent of Cheating" that occurred on the exam. To do this, we calculated the proportion of the multiple-choice exam questions that were answered (i.e., not skipped) by the student and that were categorized as instances of cheating. An Extent of Cheating of 0.50, for instance, would indicate that the student had cheated on half of the multiple-choice questions that they answered on the exam.

For each exam taken by a student, we then categorized the *exam as a whole* as an instance of "cheating" or "not cheating" based on the Extent of Cheating present on the exam. If

the Extent of Cheating was a proportion of 0.15 or greater, then that exam was scored as cheated. We chose that cutoff point to avoid potential false positives caused by a student leaving the test page once or twice for reasons other than cheating. An Extent of Cheating of 0.15 or greater would indicate that the student left the exam page for more than on 3 out of the 20 questions. As described below, this cutoff value led to extremely few borderline cases; exams categorized as instances of cheating almost universally showed Extents of Cheating far greater than 0.15.

Standardization of Test Scores

To address our research questions, we needed to make comparisons between different exams within the course, which were not necessarily of equal difficulty. To enable those comparisons, we converted students’ raw test scores to standardized ones. To do that, we first calculated the average exam score and standard deviation for exams taken during three previous semesters of running the course (all with proctored exams). We used those historical data to provide an estimate of the degree of difficulty for each exam. We then converted students’ exam scores for the summer 2020 section to Z scores based on the historical means and standard deviations.

Results

Prevalence of Cheating

Table 1 summarizes the testing conditions for each of the exams as well as the prevalence of cheating behaviors on each one. In the sections that follow, we discuss how the different testing conditions affected rates of cheating. Worth noting at the outset, however, are the very high rates of cheating that occurred during the first four exams.

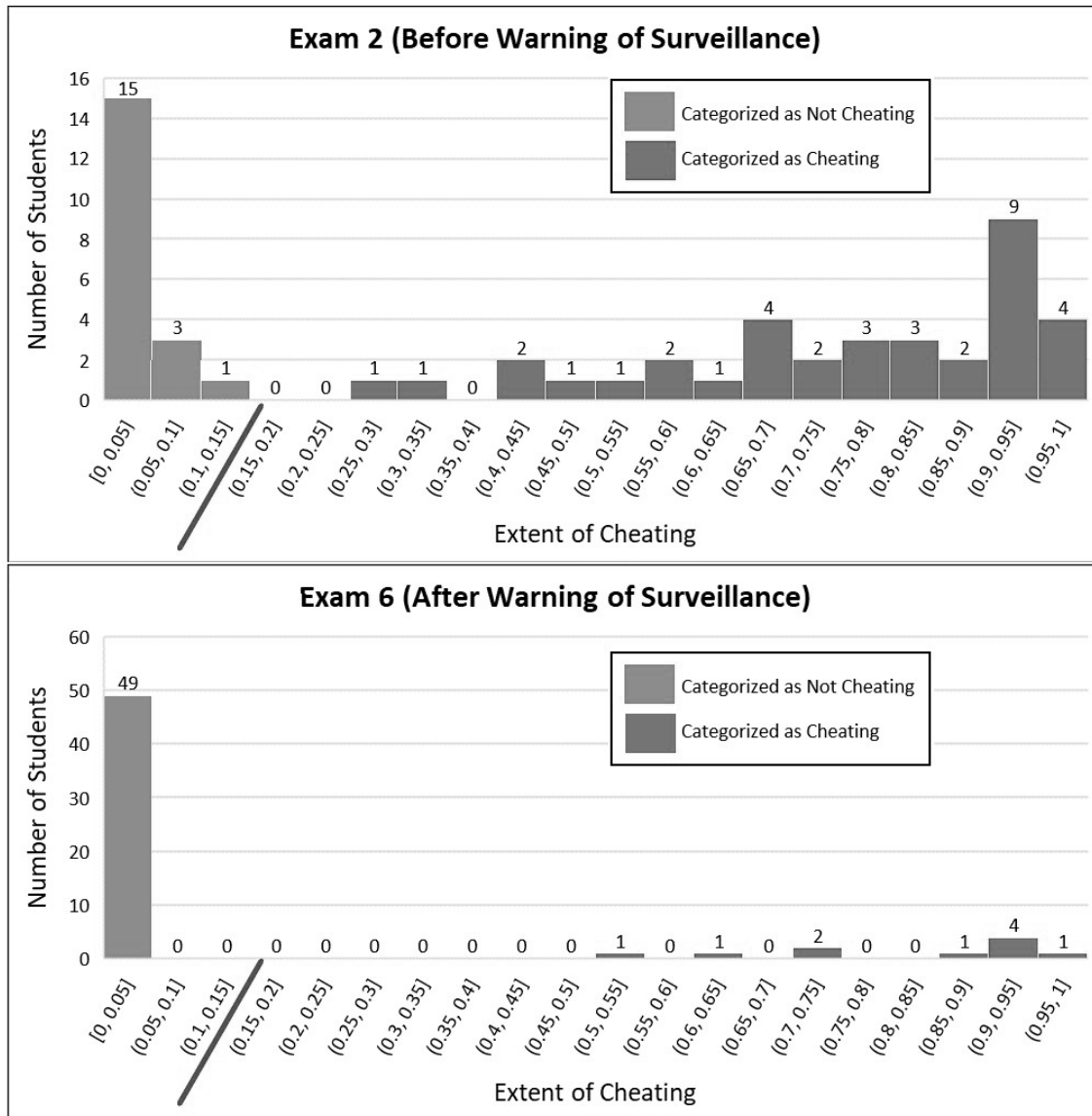
Table 1
Frequency and Extent of Cheating Behavior

Testing Conditions		n	Students Engaging in Cheating Behavior (%)	Average Extent of Cheating ^a	
Exam 1	Restricted Time	Appeal Group	31	68	0.69
		Pledge Group	30	70	0.77
Exam 2		Honesty Pledge	62	69	0.75
Exam 3	Less Restricted Time	Honesty Pledge	61	67	0.79
Exam 4				72	0.71
Exam 5	Less Restricted Time	Surveillance Warning	59	20	0.71
Exam 6				17	0.81
Exam 7	Less Restricted Time	Surveillance Warning	60	10	0.50
Exam 8		+ Emails	60	12	0.72

Extent of Cheating is defined as the proportion of multiple-choice questions on which a student cheated. Average Extent of Cheating is calculated using data only from students who engaged in cheating behavior.

On each exam, approximately 70% of students were engaged in cheating. Moreover, when students *did* cheat, they tended to do so a great deal. Figure 1 shows the distribution of the extent of cheating on Exam 2 and Exam 6, which are representative of exams with high rates of cheating and low rates of cheating, respectively. While five of the students who cheated on Exam 2 did so on less than half of the questions, most students did so on the majority of the questions, and 15 students cheated on 90% or more of the questions. Although relatively few students cheated on Exam 6, those who did showed a similar pattern in that they tended to cheat on most of the questions rather than just a few.

Figure 1
Extent of Cheating on Two Representative Online Exams



Note. Extent of Cheating is defined as the proportion of multiple-choice exam questions for which there was evidence of cheating. The cutoff point (marked by the red line) for categorization was defined as an Extent of Cheating of 0.15.

Effects of Mitigation Measures on Cheating Behaviors

Appeal versus Pledge of Honesty

As can be seen in Table 1, the rates of cheating as well as the extent of cheating were high for both the Appeal and Pledge groups. Between the two groups, there was a small apparent difference between the proportion of students who cheated and a slightly larger apparent difference in the extent of cheating. To test whether those differences are statistically significant, we first used a Z-test to compare the percentage of students who cheated across the two groups; the Z-test is appropriate here as it allows for the comparison of proportions. The results of that test indicate that the small difference between the two groups is not statistically significant ($Z = 0.169, p = .865$). To examine whether the different extent of cheating between the Appeal and Pledge conditions was statistically significant, we used an independent-samples t-test. The t-test was appropriate in this case given that we were comparing mean values (extent of cheating) rather than proportions. We found no statistically significant difference in the extent of cheating between the groups ($t(41) = 1.02, p = .318$). In sum, neither an appeal nor honesty pledge appears to be particularly effective at curbing student engagement in cheating behavior. Because we found no statistically significant differences between the two conditions, data from these two groups were combined for all the analyses that follow.

Time Limits

Table 2 provides summary statistics for the time taken on the first four exams. As a point of comparison, we also include historical exam times taken from the previous three semesters of the course. As shown in Table 2, the first two exams had a relatively tight time restriction, which was then relaxed for Exams 3 and 4. After the time limits were relaxed, there is an apparent increase in exam times for summer 2020 students. To investigate whether that increase was statistically significant, we used a paired-samples t-test to compare students' time taken on Exam 2 and Exam 3. We used a paired test here because we were comparing students' time taken on Exam 2 to their own times to complete Exam 3 (paired tests are used in many subsequent analyses for the same reason). The results of that test indicate that, on average, students took longer to complete Exam 3 than they did to complete Exam 2 ($t(64) = 5.649, p < .0001$). The increase is unlikely to be attributable to the relative lengths of the exams; as seen in the historical data, students have generally taken less time, not more, on Exam 3 versus Exam 2.

As time limits were relaxed, we investigated whether the percentage of students who cheated changed from Exam 2 to Exam 3. As shown in Table 1, there is a small apparent difference in the proportion of students who cheated on those two exams. We used a Z-test to compare those two proportions but found that the difference was not statistically significant ($Z = -0.234, p = .810$). Among students who cheated, the extent of cheating also did not significantly change when time constraints were relaxed. For students who cheated on both exams, we compared their extent of cheating on Exam 2 and Exam 3 using a paired-samples t-test but found no statistically significant difference ($t(41) = 0.723, p = .474$).

Given that students took more time on Exam 3, we wondered whether it was the students who were cheating who were using that additional time, perhaps to cheat more intensively on each question. However, we did not find that to be the case. We used a between-samples t-test to compare how much *additional* time was used on Exam 3 versus Exam 2 between those who cheated and those who did not; we found no statistically significant difference ($t(55) = 1.470, p = .147$). In sum, we have no evidence that time limits have any meaningful effect on cheating behaviors. An additional analysis of the relationship between exam times and cheating behaviors

can be found in the supplemental materials; that analysis provides further support for the results described here.

Table 2
Time Taken on Exams for Current and Past Sections

	Historical Exam Times (Proctored)				Summer 2020 Exam Times			
	n	Limit (min)	Mean (min)	SD (min)	n	Limit (min)	Mean (min)	SD (min)
Exam 1	131	30	13.68	7.17	65	20	17.82	2.69
Exam 2	129	30	13.30	7.30	65	20	17.86	3.02
Exam 3	129	30	11.92	7.33	66	30	20.54	5.98
Exam 4	132	40	17.72	9.14	66	40	23.62	7.75

Note. More restrictive time limits are noted in **bold**.

Warning of Surveillance

After the first four exams, all students were issued a warning on each of the remaining exams stating that they were being surveilled and that any dishonest behavior would result in disciplinary action. Evident in Table 1 is a large apparent reduction in cheating behavior after Exam 4, dropping from 72% on Exam 4 to 20% on Exam 5. To determine whether that reduction was statistically significant, we used McNemar's X^2 Test, which allowed us to compare the proportion of students who *changed* their behavior from Exam 4 to Exam 5. The results of that test indicate that the change in behavior was statistically significant (McNemar's $X^2 = 31.03$, $p < .0001$). Importantly, this finding provides strong evidence that the behaviors observed in the Action Logs are, in fact, indicative of cheating; no other apparent explanation exists for the sharp reduction in the behavior as a result of the warning. Interestingly, among students who continued to engage in cheating behaviors after the warning, we saw no change in the *extent* of cheating from Exam 4 to Exam 5.

After Exam 6, students still engaging in cheating behavior were sent a personal communication notifying them that their behavior had been detected and that they would not receive credit if they continued to engage in that behavior. Four students ceased engaging in cheating behaviors after receiving the personal communication following Exam 6, and another two who continued to cheat on Exam 7 ceased engaging in cheating behavior after a follow-up email. Three students continued to engage in cheating behavior through Exam 8 despite the personal warning emails. Personal communications therefore did seem to further reduce cheating behaviors but not fully extinguish it. Interestingly, three students who had previously ceased cheating after Exam 5 re-engaged in cheating on Exam 8. Although occurring only in a small number of students, this finding does raise the possibility that students might stop taking warnings of surveillance seriously over time, thus requiring personal messages to reinforce the warning.

Interactions Between Cheating Behavior and Exam Scores

Table 3 provides summary statistics for the historical test score data as well as raw and standardized scores for the summer 2020 section of the course. Test scores and standard deviations are reported as percentages of total possible points on the exam. Unless otherwise noted, all of the analyses that follow use the standardized scores rather than the raw values.

Table 3
Summary of Exam Scores

	Historical Exam Scores (Proctored)			Summer 2020 Exam Scores		
	n	Mean (%)	SD (%)	n	Mean (%)	SD (%)
Exam 1	131	66.49	17.28	65	78.85	12.01
Exam 2	129	62.25	20.29	65	74.42	14.80
Exam 3	129	63.02	19.59	66	83.41	13.88
Exam 4	132	69.33	18.21	66	82.99	12.54
Exam 5	129	64.64	19.28	63	75.47	16.02
Exam 6	127	61.54	18.14	63	70.16	19.77
Exam 7	121	68.84	19.72	64	79.52	14.87
Exam 8	123	75.81	17.57	64	81.80	11.70

	Summer 2020 Standardized Scores		
	n	Mean	SD
Exam 1	65	0.71	0.70
Exam 2	65	0.60	0.73
Exam 3	66	1.04	0.71
Exam 4	66	0.75	0.69
Exam 5	63	0.56	0.83
Exam 6	63	0.48	1.09
Exam 7	64	0.54	0.75
Exam 8	64	0.34	0.67

Note. Non-standardized exam scores and standard deviations are expressed as a percentage of possible points earned on the exam, with 100% representing the highest possible score.

Cheating was prevalent on Exams 1 to 4, and students’ scores on those exams were also higher than historical averages. Compared to past iterations of the course, students were on average scoring 0.78 standard deviation units above the historical mean for those exams. A one-sample t-test confirmed that the higher test scores were significantly higher than the historical means ($t(64) = 12.165, p < .00001$). As described above, when a warning of surveillance was issued beginning with Exam 5, we found that the prevalence of cheating declined dramatically. If cheating were responsible for the elevated test performance seen in Exams 1 to 4, then the cessation of cheating should coincide with a decline in test performance. Indeed, we did find that average test scores declined along with the prevalence of cheating. When we compared average standardized exam scores on Exams 3 and 4 versus Exams 5 and 6, a paired-samples t-test indicated a statistically significant decrease in scores ($t(61) = -3.54, p = .0008, 95\% \text{ CI for difference} = (-0.547, -0.254)$).

However, the overall changes in exam scores represent aggregate-level comparisons, and a more nuanced view can be obtained by examining differences between students who did and did not engage in cheating behaviors. We would expect to observe the reduction in exam scores primarily for students who *stopped cheating*. We would not expect a reduction in exam scores for students who *never* engaged in cheating or those who *continued* to cheat. To test that conjecture, we conducted a mixed two-way ANOVA. That ANOVA model allows one to compare how a response variable of interest (in this case, exam scores) is related to multiple interacting factors (in this case, cheating behavior as well as changes in the test conditions).

The response variable in the ANOVA model was standardized exam score for a targeted set of exams. The within-subjects factor (EXAM) had two categorical levels, corresponding to the two pairs of exams that were of interest: Exams 3 and 4 (on which students made an honesty pledge) versus Exams 5 and 6 (where students were given a warning of surveillance). We

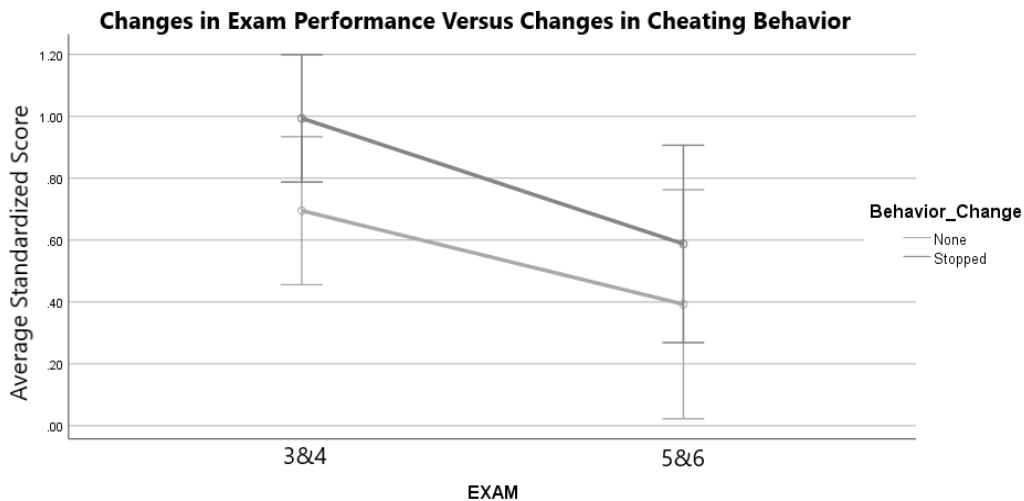
focused on pairs of exams for several reasons. First, the exams within each pair had identical testing conditions; a student’s average performance within each pair therefore provides a reasonable estimate of their performance under those conditions. Second, students were given testing deadlines for pairs of exams rather than individual ones. The possibility therefore exists that students allocated less time and effort on the second of any pair of exams due to the way that they managed their time. By averaging across exam pairs, any effect from that possibility is controlled.

The between-subjects factor (BEHAVIOR CHANGE) had two categorical levels, corresponding to whether a student showed a marked change in cheating behavior between the two pair of exams. Students were categorized as “none” for this variable if they never cheated on Exams 3 to 6 or if they cheated on all of those exams. Students were categorized as “stopped” if they had cheated on both Exams 3 and 4 but did not cheat on either Exams 5 or 6. The ANOVA analysis included data from 54 students for whom we had complete sets of behavioral and performance data. Of those students, 23 showed no change in behavior and 31 stopped cheating. No assumptions of the ANOVA model were found to be violated; a null result was found for Levene’s test for equality of error variances for mean scores on Exams 3 and 4 ($p = .162$) as well as Exams 5 and 6 ($p = .377$).

Contrary to our expectations, no statistically significant interaction was found between the BEHAVIOR factor and the EXAM factor ($F_{(1,52)} = 0.382, p = .539$, partial $\eta^2 = .007$). A large and statistically significant main effect was found for EXAM ($F_{(1,52)} = 17.903, p < .001$, partial $\eta^2 = .256$) but not for BEHAVIOR CHANGE ($F_{(1,52)} = 1.741, p = .193$, partial $\eta^2 = .032$). These results are illustrated by the interaction plot in Figure 2. In sum, they indicate that the exam scores for all students, regardless of whether they always cheated, never cheated or stopped cheating, declined similarly between Exams 3 and 4, before the warning, and Exams 5 and 6, after the warning. Because the exam scores have been standardized, that score reduction cannot be attributed to changes in exam difficulty. The reduction in scores across all students is therefore both an unexpected and puzzling result.

Figure 2

Interaction Plot for the EXAM and BEHAVIOR CHANGE factors (Error bars represent 95% confidence intervals)



Discussion

Two of our research goals were to determine the prevalence of cheating during unproctored online exams and the effects of various interventions on reducing cheating behaviors. In the absence of warnings of surveillance, we found cheating behaviors to be widespread. Neither appealing to students' academic integrity nor requiring an honesty pledge were found to be effective, as approximately 70% of students were observed cheating under either condition. It is possible that cheating was even more widespread than what we report here, as we were only able to detect a certain type of cheating behavior. In addition to detecting a high prevalence of cheating, we also found that when students did cheat, they did so on the majority of questions on a given exam rather than just one or two.

The pervasiveness of cheating during unproctored exams is sobering. Previous studies that found evidence of cheating (e.g., Alessio et al., 2017; Fask et al., 2014; Hylton et al., 2016) relied on aggregate measures and so could not estimate the prevalence of cheating. Unfortunately, our results indicate that cheating is the norm rather than the exception. One possible reason for our findings is that the type of cheating investigated here (using unauthorized sources to look up answers) is seen by students as relatively acceptable (Dyer et al., 2020). Students might not regard looking up answers as a "serious" or even "real" form of cheating, unlike other forms such as copying a peer's work or having a peer take a test in their stead. Another possible reason why consulting unauthorized materials on a computer is so common is that it is simply easy to do. Navigating away from a test page to search through notes or the internet requires little premeditation and little investment in time (we found no evidence that students who cheated took any longer on the exams). It is, in most respects, a completely natural impulse when using a computer. Our results would indicate that most students do not suppress that impulse unless they believe that their behavior is being monitored.

Whatever the reason for the pervasiveness of cheating, it is clearly a serious problem and not simply an unfounded worry. Our finding could be used to argue for the necessity of proctoring technologies, but we also found that cheating behaviors could be substantially reduced using far less invasive, costly, or cumbersome methods. Although we were unable to completely eliminate cheating behaviors, we found that warning students who continued to cheat that their cheating had been detected was highly effective in further reducing cheating.

We emphasize that we could only detect a certain type of cheating behavior. When students stopped engaging in that specific behavior, they might very well have switched to some other form of cheating that we could not detect. For instance, they could have consulted printed materials or materials on a different device. Although we cannot rule out that possibility, we think it unlikely. As noted above, we suspect that consulting unauthorized material on a computer is so common because it is both easy to do and consistent with typical computer usage. In contrast, shifting to an undetectable cheating method would require deliberate planning and preparation. Although some number of students might make the effort to cheat in those ways, we suspect that the proportion of students would be far less than the three-quarters who we detected cheating in this study. Additionally, the warning sent to students was nonspecific in that they did not know what kinds of behaviors we could and could not observe. Students can only shift to undetectable cheating methods if they know what is and is not detectable.

Warning students that they are being surveilled and that serious consequences await those who are detected cheating is effective, but we also emphasize that the warning requires follow-up. We found several instances of students who stopped engaging in cheating behaviors after

receiving the warning only to re-engage in those behaviors on later exams. We also found that a small number of students continued to cheat even after being warned and being sent follow-up emails that their behaviors had been detected. In practical terms, this means that surveillance warnings should not be bluffs. Although a bluff might curb cheating in the short run, it is not likely to yield long-term results. Of course, this requires that instructors have access to something like an Action Log that can actually detect cheating.

Complicating all of the above are our findings regarding the interaction between cheating behaviors and exam scores. An intuitive assumption regarding cheating, particularly looking up answers during an exam, is that it will lead to students earning higher scores. However, we found no evidence that cheating behaviors were associated with elevated test scores. When warnings of surveillance were issued to students, rates of cheating declined substantially, and we did find a corresponding decline in test scores. Yet our analyses revealed that the exam scores of *all* students declined, including scores for students who never cheated and those who continued to cheat.

Several possible explanations exist for our unexpected results. One is that students in the *never cheating* category were actually cheating a different way, such as by looking at printed notes or textbook. Because our warning was not specific as to the way in which we were monitoring student behavior, it is possible that those students thought they were being surveilled and stopped that behavior. Thus, their categorization as *never cheaters* might not have been accurate. More puzzling is that students who continued to cheat also showed a decline in exam scores. The fact that *never cheaters* and *always cheaters* both declined in their scores raises the possibility that the surveillance warning itself could have affected performance. A well-documented effect is that the level of nervousness of students when taking a test depresses test scores (Cassady & Johnsen, 2002). By issuing a warning to all students this may have increased the level of anxiety during test-taking, which would have lowered all students' scores. This is a possibility that has been suggested by previous researchers in relation to proctoring technologies (Hylton, Levy, & Dringus, 2016) and is one that warrants further study.

The fact that we were not able to find any link between cheating behaviors and exam performance suggests that cheating, at least of the sort examined here, might be far less effective in improving test scores than it is often assumed to be. For instance, we saw many examples where students cheated on nearly every exam question (see Figure 1), yet those students were not consistently answering every question correctly. That indicates finding the correct answer to a question may not be easily accomplished with a brief search of the lecture notes or an internet search. This phenomenon warrants further inquiry. A deeper analysis might reveal whether certain types of question are more resistant to cheating than others, or whether some students are more effective cheaters than others. Yet if cheating does not account for the higher test scores of students compared to historical means when 70% were cheating, what does? It is, of course, possible that the students in this particular study were simply atypical (perhaps higher achieving than past students). It is also possible that the testing environment at home might partially contribute to elevated performance, as previous research has suggested (Fask et al., 2014). Future studies should investigate that possibility.

If the kind of cheating examined in this study (consulting unauthorized materials during an exam) does not necessarily lead to elevated performance, is it still a behavior worthy of concern? Instructors might take some comfort in knowing that if their students cheat in this way (and, our results indicate, odds are good that they will), it will not necessarily lead to artificially inflated grades. However, the fact that this form of cheating is not terribly effective does not

make it any more ethical. Violations of stated testing procedures should be and are likely to be concerning for most instructors, regardless of how those violations affect students' overall grades.

One option that instructors have is to simply change their stated testing procedures and allow students to consult whatever materials they think would be beneficial. Another would be to use draconian surveillance technologies to more closely monitor students. Although those technologies might suppress cheating, our results indicate that less invasive approaches are also effective. Appealing to students' honesty or having them sign pledges is unlikely to change students' behavior, but if a warning is given to students that leads them to believe that they are being monitored, cheating is less likely to occur. Provided that the belief does not erode over time, we suggest that taking this approach is an effective way at reducing cheating during unproctored online exams. At the same time, we caution that surveillance measures might negatively impact student performance by provoking anxiety, which would affect *all* students, not just those who cheat. Colleges and universities should keep that caution in mind before investing time and resources into remote proctoring technologies.

Declarations

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The author(s) received approval from the ethics review board of Iowa State University, USA for this study.

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- 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), 146–161. <http://dx.doi.org/10.24059/olj.v21i1.885>
- Alessio, H. M., Malay, N., Maurer, K., Bailer, A. J., & Rubin, B. (2018). Interaction of proctoring and student major on online test performance. *International Review of Research in Open and Distributed Learning, 19*(5). <https://doi.org/10.19173/irrodl.v19i5.3698>
- 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>
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*(2), 270–295. <https://doi.org/10.1006/ceps.2001.1094>
- 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).
- Dunn, T. P., Meine, M. F., & McCarley, J. (2010). The remote proctor: An innovative technological solution for online course integrity. *International Journal of Technology, Knowledge & Society, 6*(1), 1–8.
- Dyer, J. M., Pettyjohn, H. C., & Saladin, S. (2020). Academic dishonesty and testing: How student beliefs and test setting impact decisions to cheat. *Journal of the National College Testing Association, 4*(1), 1–30.
- Fask, A., Englander, F., & Wang, Z. (2014). Do online exams facilitate cheating? An experiment designed to separate possible cheating from the effect of the online test taking environment. *Journal of Academic Ethics, 12*(2), 101–112. <https://doi.org/10.1007/s10805-014-9207-1>
- Flaherty, C. (2020, May 11). Big proctor. *Inside HigherEd*. <https://www.insidehighered.com/news/2020/05/11/online-proctoring-surg-ing-during-covid-19>
- Grajek, S. (2020, April 10). EDUCAUSE COVID-19 QuickPoll Results: Grading and Proctoring. *EDUCAUSE Review*. <https://er.educause.edu/blogs/2020/4/educause-covid-19-quickpoll-results-grading-and-proctoring#fn2>
- Grijalva, T., Nowell, C., & Kerkvliet, J. (2006). Academic honesty and online courses. *College Student Journal, 40*(1), 180–185.
- Harmon, O. R., & Lambrinos, J. (2008). Are online exams an invitation to cheat? *Journal of Economic Education, 39*(2), 116–125. <https://doi.org/10.3200/JECE.39.2.116-125>
- Hollister, K. K., & Berenson, M. L. (2009). Proctored versus unproctored online exams: Studying the impact of exam environment on student performance. *Decision Sciences Journal of Innovative Education, 7*(1), 271–294. <https://doi.org/10.1111/j.1540-4609.2008.00220.x>
- Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education, 92*, 53–63. <https://doi.org/10.1016/j.compedu.2015.10.002>
- Instructure. (2020). *Canvas Learning Management System* [software]. <https://www.instructure.com/canvas/>

- Kerkvliet, J., & Sigmund, C. L. (1999). Can we control cheating in the classroom? *Journal of Economic Education*, 30(4), 331–351.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644.
<https://doi.org/10.1509/jmkr.45.6.633>
- McGee, P. (2013). Supporting academic honesty in online courses. *Journal of Educators Online*, 10(1), 1–31. <https://www.learntechlib.org/p/114388/>
- McNabb, L., & Olmstead, A. (2009). Communities of integrity in online courses: Faculty member beliefs and strategies. *Journal of Online Learning and Teaching*, 5(2), 208–23.
- Respondus. (2020a). *Lockdown Browser* [software].
<https://web.respondus.com/he/lockdownbrowser/>
- Respondus. (2020b). *Respondus Monitor* [software]. <https://web.respondus.com/he/monitor/>
- Stephens, J. M., Young, M. F., & Calabrese, T. (2007). Does moral judgment go offline when students are online? A comparative analysis of undergraduates' beliefs and behaviors related to conventional and digital cheating. *Ethics & Behavior*, 17(3), 233–54.
<https://doi.org/10.1080/10508420701519197>
- Stogner, J., Miller, B., & Marcum, C. (2013). Learning to e-cheat: A criminological test of internet facilitated academic cheating. *Journal of Criminal Justice Education*, 24(2), 175–199. <https://doi.org/10.1080/10511253.2012.693516>
- Stuber-McEwen, D., Wiseley, 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(3), 1–10.
- Watson, G. R., & 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.
- Wiley. (2020, July 22). *Academic integrity in the age of online learning*.
<https://read.uberflip.com/i/1272071-academic-integrity-in-the-age-of-online-learning/0?>