Comparing the Security of Three Proctoring Regimens for Bring-Your-Own-Device Exams

Rishi Gulati
rishisg2@illinois.edu
University of Illinois, Urbana-Champaign

Matthew West
mwest@illinois.edu
University of Illinois, Urbana-Champaign

Craig Zilles
zilles@illinois.edu
University of Illinois, Urbana-Champaign

Mariana Silva
mfsilva@illinois.edu
University of Illinois, Urbana-Champaign

ABSTRACT

We compare the exam security of three proctoring regimens of Bring-Your-Own-Device, synchronous, computer-based exams in a computer science class: online un-proctored, online proctored via Zoom, and in-person proctored. We performed two randomized crossover experiments to compare these proctoring regimens. The first study measured the score advantage students receive while taking un-proctored online exams over Zoom-proctored online exams. The second study measured the score advantage of students taking Zoom-proctored online exams over in-person proctored exams. In both studies, students took six 50-minute exams using their own devices, which included two coding questions and 8–10 non-coding questions.

We find that students score 2.3% higher on non-coding questions when taking exams in the un-proctored format compared to Zoom proctoring. No statistically significant advantage was found for the coding questions. While most of the non-coding questions had randomization such that students got different versions, for the few questions where all students received the same exact version, the score advantage escalated to 5.2%. From the second study, we find no statistically significant difference between students' performance on Zoom-proctored vs. in-person proctored exams. With this, we recommend educators incorporate some form of proctoring along with question randomization to mitigate cheating concerns in BYOD exams.

CCS CONCEPTS

• **Social and professional topics → Computing education**; **Student assessment**.

KEYWORDS

bring-your-own-device, BYOD, online, proctoring, cheating, exam security, randomization, computer-based testing

SIGCSE 2024, March 20–23, 2024, Portland, OR, USA

ACM ISBN 979-8-4007-0423-9/24/03…$15.00
https://doi.org/10.1145/3626252.3630809

ACM Reference Format:
Rishi Gulati, Matthew West, Craig Zilles, and Mariana Silva. 2024. Comparing the Security of Three Proctoring Regimens for Bring-Your-Own-Device Exams. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024), March 20–23, 2024, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3626252.3630809

1 INTRODUCTION

The increasing enrollment in computer science (CS) courses calls for effective strategies to meet this demand and to efficiently conduct large-scale exams [10, 16, 26]. Traditional pen-paper exams present logistical hurdles such as requesting space, printing exams, proctoring costs, and timely grading and feedback [38, 50]. To circumvent these challenges, computer science faculty are actively turning to computer-based exams [14, 40] and automation (e.g., auto-grading) for managing this growth [8, 24, 27]. Such computer-based exams reduce the resources required for grading, offer faster feedback, and provide a more authentic testing environment for CS exams (via the use of compilers and debuggers) [4, 20, 49].

The recent ubiquity of mobile computing devices like laptops and tablets (especially among CS majors) has inspired a Bring-Your-Own-Device (BYOD, originally coined by Ballagas et al. [2]) model that enables running computer-based exams without large investments in infrastructure [9, 13]. In a typical BYOD model, the students use their own personal laptops for all classroom learning activities, including both high- and low-stakes assessments. Just like any new advancement, the initial adoption of BYOD examinations was slow. From the perspective of students, the concerns revolved around privacy and confidentiality of personal data in their devices [1, 41, 48]. However, in more recent times, a renewed interest in BYOD assessments has emerged, backed by a series of academic pursuits underscoring its many merits like academic achievement [42], quality of work and in-class motivation [22, 42], anxiety reduction [47], and student perception and satisfaction [25, 36, 43]. Student awareness and preference for BYOD exams have further paved the way for its adoption by institutions.

However, there are still challenges associated with BYOD exams. From the perspective of educators, these are centered around exam integrity and the perception of the ease of cheating [6, 15, 21]. As postulated by Dawson et al., "The BYOD eExam is by definition less secure than both pen-and-paper examinations and examinations held in a computer laboratory, as it has all the vulnerabilities of both environments, as well as some of its own" [23].

The most prevalent solution for maintaining academic integrity on online/BYOD exams has been the use of lockdown browsers. There exists a corpus of research probing the efficacy of such browsers and/or specialized software that transforms student devices into secure workstations to deter dishonest practices [32, 33, 39]. Nonetheless, these mechanisms, while promising, fall short of being a reliable remedy. They struggle with securing the testing environment [37, 45], usability and cross-platform support [34], and poor student perception [1, 3, 30]. Owing to such problems, negative student sentiment, technical support issues, and a strong pedagogical desire to run "open-notes" exams, we opted to forgo the use of lockdown browsers in our studies.

Instead, we delve into an exploration of three distinct proctoring regimens: un-proctored, Zoom-proctored, and in-person proctored exams. Prior literature points out that un-proctored quizzes often lead to grade inflation [11, 39, 46] and a higher variance in student's scores [29]. Intriguingly, the magnitude of this inflation is observably amplified as the academic term advances [18]. We are interested in researching the difference in student performance between un-proctored and proctored (Zoom) BYOD examinations for CS courses. Another crucial undertaking of our research is to tease out the difference in performance dynamics between proctoring environments with respect to the question type (coding vs non-coding).

While in-person proctoring is logistically difficult, it has a long history and many instructors are comfortable with it. Yet, emerging literature implies that transitioning from traditional in-person proctoring to a Zoom-based proctoring system exerts negligible impact on student performance metrics [17, 28, 35]. This raises questions regarding the applicability of such findings to BYOD-oriented computer science examinations. Our interests lie in discerning whether computer science students, who are typically more tech-savvy than students in general, have an increased potential for exploiting the BYOD mode of assessment.

We also consider the potential enhancements to exam security by integrating various proctoring methodologies with question randomization techniques. Historically, the utility of question randomization has been well-documented in the domain of computer-based assessments, with a notable body of literature emphasizing its efficacy [31, 44, 51]. Introducing variability in the examination content inherently discourages conventional cheating tactics, with Chen et al. [19] finding that educators only need 3–4 different questions in a pool, each with randomized parameters, to significantly curtail cheating. However, the context of a BYOD environment presents a novel challenge. If students are already familiar with the intricacies of their own devices and the myriad ways in which they can access external information, is question randomization still effective as a cheating deterrent?

In this paper, we examine BYOD exams under three proctoring methodologies coupled with question randomization as potential deterrents to academic misconduct. Specifically, we address the following research questions:

RQ1: Do students score higher on BYOD exams when they are un-proctored, relative to Zoom proctored? If yes, then to what extent?

RQ2: Do students score higher on BYOD exams when they are Zoom-proctored relative to in-person proctored?

RQ3: Does the use of randomized problems help mitigate cheating in BYOD exams?

2 METHODS

The studies were conducted at a large public research university in the United States during the Fall 2021 and Spring 2022 semesters, focusing on a required upper-division computer science course with students expected to have prior programming experience through a CS1 course prerequisite. For both studies, six BYOD 50-minute exams were administered, each accounting for 7% of the final grade. The exams included two coding questions and 8-10 non-coding questions (numeric-input, multiple-choice, checkbox), delivered through an online tool allowing for parameterized question instances, ensuring non-identical exams. The exams were autograded with immediate feedback, permitting retries for reduced credit on non-coding questions and unlimited attempts on coding questions. Students were allowed access to any web or device content (including IDEs), with the caveat that they were not allowed to communicate with others during the exam.

For Zoom proctoring, students needed a second device, such as a phone or tablet, to capture their face, computer screen, and work area on Zoom. This device was solely for proctoring purposes, with students completing the exam on their primary device, in which they were not connected to Zoom. The Zoom sections maintained a ratio of about 45 students to one proctor. In un-proctored or in-person exams, students accessed the exam through the same system but without Zoom. In-person exams occurred in the classroom with a 20-to-1 student-to-proctor ratio.

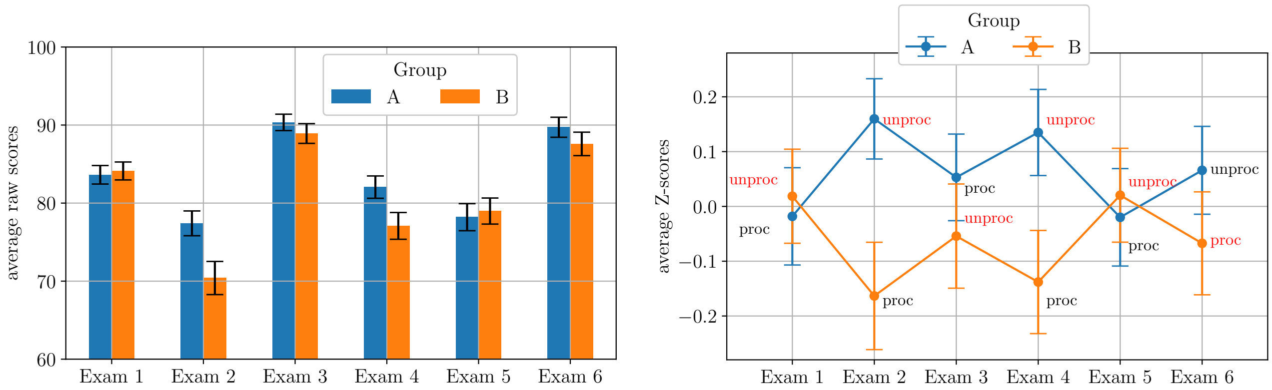
The two studies were conducted using a randomized cross-over design. Study 1, in Fall 2021, included 374 students randomly split into groups A and B, alternating between un-proctored online and Zoom-proctored online exams (Table 1). After filtering to only students who completed all exams without test accommodations, 263 students remained: 7% freshmen, 34% sophomores, 38% juniors, and 21% seniors.

Table 1: Exam schedule for groups A and B (Study 1).

	Exam 1	Exam 2	Exam 3	Exam 4	Exam 5	Exam 6
A	proc	unproc	proc	unproc	proc	unproc
B	unproc	proc	unproc	proc	unproc	proc

Students were divided into two groups stratified by gender. After filtering, group A consisted of 79% male (105 students) and 21% female (28 students), and group B was 77% male (100 students) and 23% female (30 students). Both groups had similar major distributions, with CS being the predominant major for both group A (44%, 58 students) and group B (42%, 54 students). The average incoming Grade Point Averages (GPAs) were not statistically significantly different, with group A at 3.71 and group B at 3.75.

In the second study (Spring 2022), conducted in the in-person section of the course with 125 registered students, we followed a similar pattern, except that this time the groups alternated between Zoom-proctored online exams and in-person proctoring (Table 2). The student breakdown was 19% freshmen, 60% sophomores, 15%



(a) Raw scores

(b) Z-scores

Figure 1: Average raw (a) and standardized (b) quiz scores for groups A and B (Study 1). The error bars are the 95% confidence intervals. The group's treatment is indicated by the labels 'proc' and 'unproc' in (b).

juniors, and 6% seniors. Covid protocols forced all students to take Exam 1 online, affecting the intended setup. Again, stratification by gender was performed on the 125 students, using data only from 99 students who completed all exams without testing accommodations. After filtering, group A was composed of 80% male (39 students) and 20% female (10 students), while group B was 82% male (41 students) and 18% female (9 students). Similar to Study 1, CS was the majority major for Group A (63%, 31 students) and Group B (48%, 24 students). The average GPAs were not statistically significantly different, with group A at 3.67 and group B at 3.78.

Table 2: Exam schedule for groups A and B (Study 2).

	Exam 1	Exam 2	Exam 3	Exam 4	Exam 5	Exam 6
A	online	class	online	class	online	class
B	online	online	class	online	class	online

The collection of course data, analysis, and publication in an aggregated anonymized form was approved by the institution's IRB. Additionally, online informed consent was collected from students at the beginning of the semester.

3 RESULTS

3.1 Proctored vs. un-proctored Results (Study 1)

This section is further divided into three subsections.

3.1.1 Overall Course Analysis.

The raw average scores for groups A and B are shown in Fig. 1a, but the effects of the two treatments (un-proctored online vs. proctored online) are more clear in Fig. 1b, which illustrates the average z-scores for each group. Two main insights are apparent: Group A generally performs better, and both groups do relatively better in the un-proctored setting, as indicated by the complementary sawtooth patterns.

For inspecting these trends, we fit an ordinary least squares (OLS) model that is suitable for quantifying the relationship between the proctoring method and exam scores while controlling confounding variables like GPA.

$$y_{ij} = \sigma_j + \alpha\, GPA_i + \beta \mathrm{A}_{ij} \tag{1}$$

where the left-hand-side value y_{ij} is the raw exam score that student i received in exam j and A_{ij} is an indicator variable that is 1 if student i took the exam j in an un-proctored manner, otherwise A_{ij} is 0. We used GPA as control variable, where GPA_i is the incoming GPA of student i. Variables β, α and σ_j are the regression parameters that we want to estimate and can be interpreted as follows:

- σ_j: The mean score of exam j
- α: The coefficient corresponding to the ability of student i
- β: The score advantage for students taking an exam in an un-proctored setting (the value that we're seeking)

The regression model shows that students gain a 2.3 percentage point advantage in un-proctored settings over Zoom-proctored ones (β = 2.326, 95% CI [0.66, 3.99], p = 0.006). Though modest compared to exam averages, this increase is statistically significant. Using standardized z-scores in a similar analysis, the advantage is found to be 0.116 standard deviations, a small effect size (β = 0.116, 95% CI [0.02,0.49], p = 0.01).

3.1.2 Coding vs. Non-coding Questions.

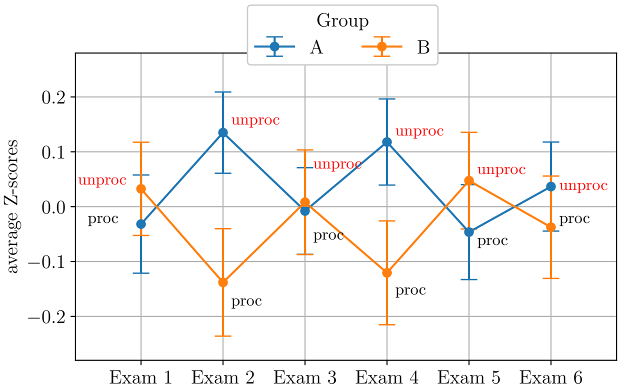
Coding questions typically test application and problem-solving skills, while non-coding questions often assess theoretical understanding and knowledge recall. This distinction is particularly relevant in CS, and could potentially affect student's cheating tendencies and strategies.

Our results show that proctoring impacts coding and non-coding questions differently. Raw scores reveal that group A performs better overall, and coding questions score lower on average (Table 3). The influence of proctoring is visible in the standardized scores for both question types across groups in Fig. 2. The sawtooth pattern, more pronounced in non-coding questions, indicates an advantage in un-proctored assessments. This pattern is less significant in coding questions.

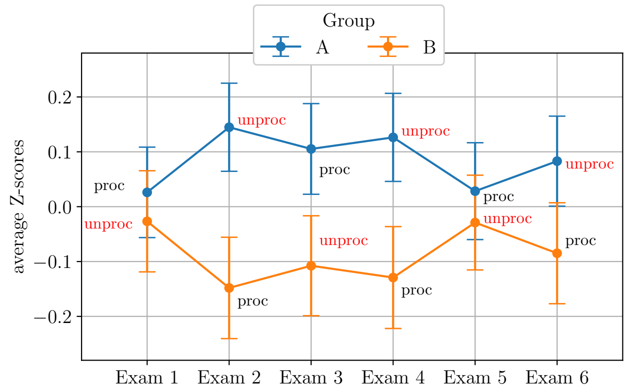
For each exam, we fitted two ordinary least squares (OLS) models:

$$\mathrm{yn}_{ij} = \sigma_j + \alpha\, GPA_i + \beta_n \mathrm{A}_{ij} \tag{2}$$

$$\mathrm{yc}_{ij} = \sigma_j + \alpha\, GPA_i + \beta_c \mathrm{A}_{ij} \tag{3}$$



(a) Non-coding questions



(b) Coding questions

Figure 2: Average z-scores for groups A and B (Study 1). The error bars are the 95% confidence intervals. The group's treatment is indicated by the labels 'proc' and 'unproc'.

Table 3: Average scores with respect to question type.

Group	Non-coding questions	Coding questions
A	86.78	69.02
B	85.31	63.68

Table 4: Regression parameters for Equations (2) **and** (3)

Parameter	coefficient	p-value	95% CI
β_n	2.39	0.004	[0.76, 4.02]
β_c	2.30	0.11	[-0.54, 5.16]

where the left-hand-side value yn_{ij} is the raw exam score that student i received in exam j on non-coding questions, and yc_{ij} is the raw exam score that student i received in exam j on coding questions. The parameters β_n and β_c and A_{ij}, α, and σ_j have similar roles to the model in Equation (1).

The results in Table 4 show that the score advantages on both question types are similar (2.39 vs 2.30 percentage points), but only the non-coding advantage is statistically significant. Repeating the analysis with standardized z-scores shows a 0.13 standard deviation advantage in non-coding questions ($\beta = 0.126$, $p = 0.01$). The effect on coding questions is not statistically significant and would be smaller due to higher score variance for this question type.

3.1.3 *Question Randomization.*

The exams under study utilized two randomization methods to deter cheating by creating slightly varied exams. First, questions were created using random parameters, and second, the exams were constructed using pools of similar problems, with students receiving random draws. In this study, a few questions had no randomization, with one or two non-coding questions being identical for all students (only in exams 1-5). Selective randomization was used to evaluate the effectiveness of parameter randomization in preventing cheating on BYOD exams.

The overall results in Table 5 reveal that the score difference between proctoring methods is more substantial (5.4 vs. 2.1 percentage points) on non-randomized questions. The z-scores for both non-randomized (Fig. 3a) and randomized (Fig. 3b) questions show a sawtooth pattern indicating an advantage in un-proctored scores, more pronounced for non-randomized questions.

Table 5: Average scores of students with respect to proctoring regimen and question randomization and their percentage point difference.

Treatment	Proctored	Un-proctored	Δ
Randomized	75.5	77.6	2.1
Non-randomized	73.3	78.7	5.4

Table 6: Regression parameters for Equations (4) **and** (5)

Parameter	coefficient	p-value	95% CI
β_n	5.19	0.016	[0.959, 9.426]
β_r	2.09	0.023	[0.284, 3.90]

To quantify the score advantage after controlling for incoming GPA, we fit two OLS regression models:

$$\mathrm{yn}_{ij} = \alpha\, GPA_i + \beta_n \mathrm{A}_{ij} \quad (4)$$

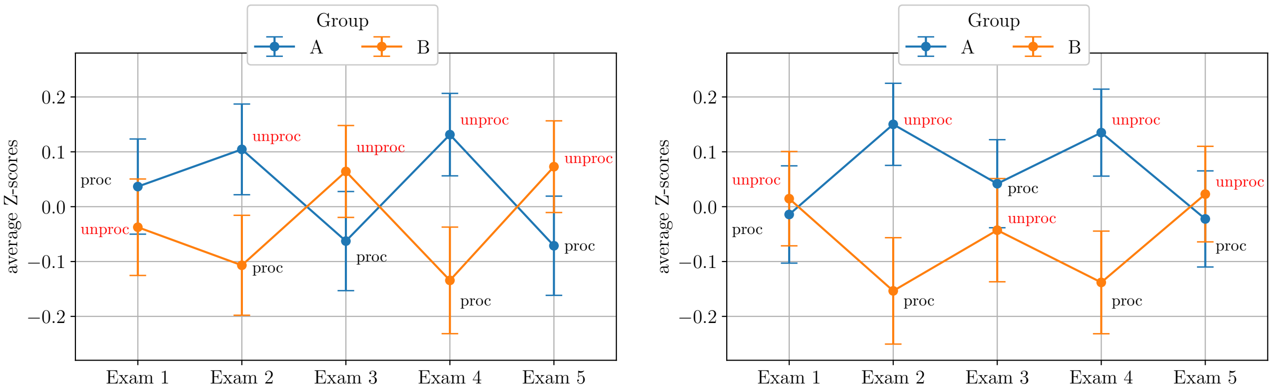
$$\mathrm{yr}_{ij} = \alpha\, GPA_i + \beta_r \mathrm{A}_{ij} \quad (5)$$

where yn_{ij} and yr_{ij} are the raw exam scores that student i received in exam j on non-randomized and randomized questions, respectively. Similarly, β_n and β_r, A_{ij}, α, and σ_j have similar roles to the model in Equation (1).

Table 6 indicates the un-proctored score advantage on non-randomized questions is 5.2 percentage points, compared to 2.1 for randomized questions, both statistically significant. When analyzed with standardized z-scores, the effect sizes are 0.13 ($\beta = 0.132$, $p = 0.015$) for non-randomized and 0.11 ($\beta = 0.109$, $p = 0.041$) for randomized questions, both considered small effect sizes.

3.2 Zoom vs. In-person Results (Study 2)

This section compares Zoom proctoring to in-person proctoring. Raw scores are shown in Fig. 4a, with group B scoring higher on 5 of the 6 exams. The comparison between Zoom and in-person



(a) Non-randomized questions **(b) Randomized questions**

Figure 3: Average z-scores for groups A and B (Study 1). The error bars are the 95% confidence intervals. The group's treatment is indicated by the labels 'proc' and 'unproc'.

proctoring does not reveal a clear trend. Figure 4b shows standardized averages, suggesting some exams may slightly favor in-person proctoring, but there is no consistent advantage.

To see if there was a statistically significant difference after controlling for incoming GPA, we fit the OLS regression model:

$$y_{ij} = \sigma_j + \alpha\, GPA_i + \beta A_{ij} \tag{6}$$

where y_{ij} is the raw exam score that student i received in exam j, A_{ij} is an indicator variable that is 1 if student i took the exam j in a Zoom-proctored manner, otherwise A_{ij} is 0 for in-person proctoring, and GPA_i is the incoming GPA of student i. The parameters α, β, and σ_j have similar roles to the model in Equation (1).

From the regression model, we find that the students score 1.5 percentage points lower when taking a BYOD exam in a Zoom-proctored setting over an in-person proctored setting (β = -1.47, 95% CI [-4.18, 1.23], p = 0.285). However, this score disadvantage is not statistically significantly different from zero. Because the overall results were not statistically significant, we did not explore finer granularity differences.

4 DISCUSSION

RQ1. Do students score higher on BYOD exams when they are un-proctored, relative to Zoom proctored? If yes, then to what extent?

Study 1 identified a significant 2.3 percentage point score advantage for students in un-proctored exams compared to Zoom proctoring. This advantage was significant for non-coding questions, but not for coding ones.

Although we do not have direct evidence, we strongly suspect that increased cheating on un-proctored exams explains this difference. Two caveats are noted. First, prior work has found that test anxiety negatively impacts students' performance on exams [5, 7, 12]. It is possible that some of the difference may be explained by students feeling more relaxed in an unproctored environment which mitigated the negative impacts of test anxiety on performance. Second, the 2.3 percentage point advantage represents only the extra cheating enabled by removing proctoring, not the total cheating. It's viewed as a lower limit of cheating on un-proctored exams and the expected cheating reduction by adding proctoring.

While 2.3 percentage points may sound to some like a relatively modest amount of cheating, it is important to recognize that this score advantage is likely not uniformly distributed among the students. If, for example, two-thirds of the class was honest and didn't attempt cheating (and thus had no score advantage), then the cheating third of the class would have an almost 7 percentage point advantage on the exam, which is approaching a full letter grade.

RQ2. Do students score higher on BYOD exams when they are Zoom-proctored relative to in-person proctored?

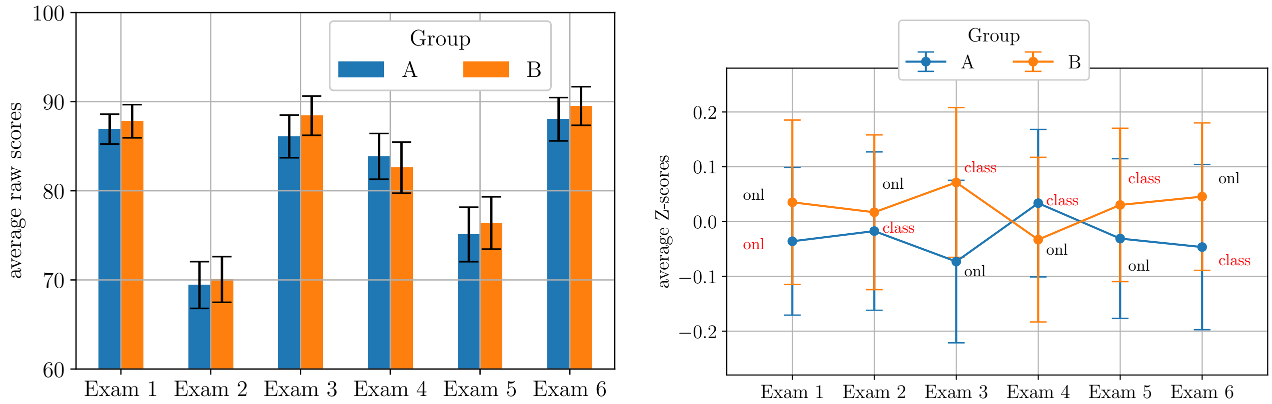
Study 2 found no significant difference between Zoom and in-person proctoring. The lack of statistical significance might be due to fewer participants, but there's also no consistent trend in the data. Contrary to expectations, regression suggests students performed better with in-class proctoring, even though Zoom proctoring seems to offer more opportunities for cheating. In addition, one might hypothesize that students would be more comfortable taking exams in their personal spaces (compared to a lecture hall), but, whether they are or not, we find no support in the data to confirm it.

Overall, we find the result that Zoom proctoring appears to have similar security to in-person proctoring to be exciting. Zoom proctoring can eliminate some of the burden of assessment for both students and faculty by eliminating the need for transit and reserving space. This finding may help faculty accept online proctoring as an alternative to BYOD exams, further facilitating their use.

RQ3. Does the use of randomized problems help mitigate cheating in BYOD exams?

Our data suggests that randomization can partially mitigate cheating in un-proctored contexts. For problems with no randomization, students had a 5.2 percentage point advantage on average compared to a 2.1 percentage point advantage with randomization, with both advantages statistically significantly different from zero.

While this finding makes sense at face value and is consistent with previous work [44], it is important again to unpack the subtleties of what our experiment measured. The difference between



(a) Raw scores

(b) Z-scores

Figure 4: Average quiz scores for groups A and B (Study 2). The error bars are the 95% confidence intervals. The group's treatment is indicated by the labels 'onl' for Zoom and 'class' for in-person proctoring in (b).

these two score advantages isn't how much cheating randomization mitigates, but instead, the incremental cheating in a switch from Zoom proctoring to no proctoring that randomization mitigates. As such, we strongly expect this to be a lower bound for the amount of cheating mitigated, because the randomization likely also mitigates cheating in Zoom proctoring relative to a situation where no student communication is possible.

5 LIMITATIONS

A limitation of our study is that it can't generalize to BYOD exams that use a lockdown browser. The faculty in charge of the class that we studied elected to not use a lockdown browser because of a pedagogical desire to run "open notes" exams and to not deal with negative student sentiment towards lockdown browsers. As lockdown browsers prevent electronic communication, we expect the amount and the mechanism of cheating to change. As such, repeating this kind of study with lockdown browsers is important for future work.

In addition, since this data was collected, the availability and student awareness of powerful AI agents like ChatGPT and GitHub Copilot have skyrocketed. When this data was collected, we suspect that a primary mechanism for cheating was communication between students in the course, and leaked answers on websites such as Chegg and Course Hero. Were we to repeat the study today, we suspect that these AI agents would be a primary mechanism for cheating, both because of their effectiveness and because they do not require cooperation between students. Furthermore, with how effectively these agents are at generating code, we suspect that our results for the coding questions might change significantly.

In the realm of academic ethics, our findings suggest a modest change in scores across different proctoring methods. If the disparities are larger, one might have to make adjustments to preserve the fairness of the assessment process.

Lastly, acknowledging the study's context, our data was collected in one course at a highly selective research university in the United States. It's crucial to consider that in regions where final exams constitute the entirety of a student's grade, the incentive to cheat may be significantly amplified, thereby potentially affecting the generalizability of our findings. Additionally, even within the same institution, the results might differ for lower-division courses where a majority of students are freshmen lacking prerequisite programming knowledge. Similar experiments in other contexts would be valuable for completely characterizing this space.

6 CONCLUSION

In this paper, we perform, to our knowledge, the first controlled crossover experiment to (1) compare the security of three different proctoring regimens (un-proctored, Zoom proctoring, and in-person proctoring) for bring-your-own-device (BYOD) computer-based exams and (2) investigate the influence of question type, question randomization and proctoring on cheating in BYOD exams.

Based on our findings, we advocate the use of some form of proctoring over no proctoring at all. We find that students score 2.3 percentage points higher on average when un-proctored relative to Zoom proctored and that students perform similarly under Zoom proctoring and in-person proctoring. While the effect size of cheating in un-proctored exams is "modest" (0.116), we urge faculty to not discount it, as it is likely not uniformly distributed. Our experience suggests that cheating attempts are concentrated in a subset of students, for whom it would have a larger effect size.

In addition, we find additional evidence that question randomization is effective in deterring cheating. We find that the score advantage more than doubles (5.2 vs. 2.1 percentage points) for non-coding questions when all students receive the same version of the question. We suggest that faculty should use question randomization in any context where they are not 100% confident that cheating is negligible.

REFERENCES

[1] David G Balash, Dongkun Kim, Darika Shaibekova, Rahel A Fainchtein, Micah Sherr, and Adam J Aviv. 2021. Examining the examiners: Students' privacy and security perceptions of online proctoring services. In *Seventeenth symposium on usable privacy and security (SOUPS 2021)*. 633–652.

[2] R. Ballagas, M. Rohs, J. G. Sheridan, and J. Borchers. 2004. Byod: Bring your own device. *In Proceedings of the Workshop on Ubiquitous Display Environments*

(UBICOMP) (2004).

[3] Zahera Baniamer and Bishr Muhamed. 2022. Cheating in Online Exams: Motives, Methods and Ways of Preventing from the Perceptions of Business Students in Bahrain. In *Technologies, Artificial Intelligence and the Future of Learning Post-COVID-19: The Crucial Role of International Accreditation*. Springer, 267–282.

[4] João Paulo Barros, Luís Estevens, Rui Dias, Rui Pais, and Elisabete Soeiro. 2003. Using lab exams to ensure programming practice in an introductory programming course. *ACM SIGCSE Bulletin* 35, 3 (2003), 16–20.

[5] Jennifer Barrows, Samantha Dunn, and Carrie A Lloyd. 2013. Anxiety, self-efficacy, and college exam grades. *Universal Journal of Educational Research* 1, 3 (2013), 204–208.

[6] Razan Bawarith, Abdullah Basuhail, Anas Fattouh, and Shehab Gamalel-Din. 2017. E-exam cheating detection system. *International Journal of Advanced Computer Science and Applications* 8, 4 (2017).

[7] Hefer Bembenutty, Wilbert J McKeachie, Stuart A Karabenick, and Yi-Guang Lin. 1998. The Relationship between Test Anxiety and Self-Regulation on Students' Motivation and Learning. (1998).

[8] Luciana Benotti, Federico Aloi, Franco Bulgarelli, and Marcos J Gomez. 2018. The effect of a web-based coding tool with automatic feedback on students' performance and perceptions. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 2–7.

[9] Daniel Biella, Steffi Engert, and Dieter Huth. 2009. Design and delivery of an e-assessment solution at the University of Duisburg-Essen. In *Proceedings EUNIS*.

[10] Betsy Bizot and Stu Zweben. 2019. Generation CS, three years later. https://cra.org/generation-cs-three-years-later/ (visited August 2019).

[11] Sara Brallier and Linda Palm. 2015. Proctored and Unproctored Test Performance. *International Journal of Teaching and Learning in Higher Education* 27, 2 (2015), 221–226.

[12] Monroe A Bruch. 1981. Relationship of test-taking strategies to test anxiety and performance: Toward a task analysis of examination behavior. *Cognitive Therapy and Research* 5 (1981), 41–56.

[13] Jens Bücking. 2010. eKlausuren im Testcenter der Universität Bremen: Ein Praxisbericht. *Retrieved März* 1 (2010), 2017.

[14] Alan C Bugbee Jr. 1996. The equivalence of paper-and-pencil and computer-based testing. *Journal of research on computing in education* 28, 3 (1996), 282–299.

[15] K. Butler-Henderson and J. Crawford. 2020. A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education* 159 (2020), 104024.

[16] Tracy Camp, W Richards Adrion, Betsy Bizot, Susan Davidson, Mary Hall, Susanne Hambrusch, Ellen Walker, and Stuart Zweben. 2017. Generation CS: the mixed news on diversity and the enrollment surge. *ACM Inroads* 8, 3 (2017), 36–42.

[17] Jason CK Chan and Dahwi Ahn. 2023. Unproctored online exams provide meaningful assessment of student learning. *Proceedings of the National Academy of Sciences* 120, 31 (2023), e2302020120.

[18] Binglin Chen, Sushmita Azad, Max Fowler, Matthew West, and Craig Zilles. 2020. Learning to cheat: quantifying changes in score advantage of unproctored assessments over time. In *Proceedings of the Seventh ACM Conference on Learning@ Scale*. 197–206.

[19] Binglin Chen, Matthew West, and Craig Zilles. 2018. How much randomization is needed to deter collaborative cheating on asynchronous exams?. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. 1–10.

[20] Yuting W Chen. 2019. Work in progress: A balancing act-Evolution of assessments in an introductory programming course in ECE after curriculum redesign. In *ASEE Annual Conference and Exposition, Conference Proceedings*.

[21] Aparna Chirumamilla, Guttorm Sindre, and Anh Nguyen-Duc. 2020. Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway. *Assessment & Evaluation in Higher Education* 45, 7 (2020), 940–957.

[22] Pao-Nan Chou, Chi-Cheng Chang, and Ching-Hsin Lin. 2017. BYOD or not: A comparison of two assessment strategies for student learning. *Computers in Human Behavior* 74 (2017), 63–71.

[23] Phillip Dawson. 2016. Five ways to hack and cheat with bring-your-own-device electronic examinations. *British Journal of Educational Technology* 47, 4 (2016), 592–600.

[24] Margaret Ellis, Clifford A Shaffer, and Stephen H Edwards. 2019. Approaches for coordinating eTextbooks, online programming practice, automated grading, and more into one course. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 126–132.

[25] Ridi Ferdiana and Obert Hoseanto. 2018. The implementation of computer based test on BYOD and cloud computing environment. *International Journal of Advanced Computer Science and Applications* 9, 8 (2018).

[26] Daniel T. Fokum, Daniel N. Coore, Eyton Ferguson, Gunjan Mansingh, and Carl Beckford. 2019. Student Performance in Computing Courses in the Face of Growing Enrollments. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 43–48. https://doi.org/10.1145/3287324.3287354

[27] Georgiana Haldeman, Andrew Tjang, Monica Babeş-Vroman, Stephen Bartos, Jay Shah, Danielle Yucht, and Thu D Nguyen. 2018. Providing meaningful feedback for autograding of programming assignments. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 278–283.

[28] Elizabeth A Hall, Christina Spivey, Hailey Kendrex, and Dawn E Havrda. 2021. Effects of remote proctoring on composite examination performance among doctor of pharmacy students. *American Journal of Pharmaceutical Education* 85, 8 (2021).

[29] Kimberly K Hollister and Mark L Berenson. 2009. Proctored versus unproctored online exams: Studying the impact of exam environment on student performance. *Decision Sciences Journal of Innovative Education* 7, 1 (2009), 271–294.

[30] Elham T. Hussain, Sumya Daoud, Hussam Alrabaiah, and Amjad K. Owais. 2020. Students' Perception of Online Assessment During the COVID-19 Pandemic: The Case of Undergraduate Students in the UAE. In *2020 21st International Arab Conference on Information Technology (ACIT)*. 1–6.

[31] Beesung Kam. 2021. Relationship Between Variables of pre-and post-Randomized Questions in Computer-based Tests. (2021).

[32] Bastian Küppers and Ulrik Schroeder. 2021. FLEX: A BYOD Approach to Electronic Examinations. *Workgroups eAssessment: Planning, Implementing and Analysing Frameworks* (2021), 145–179.

[33] Oka Kurniawan, Norman Tiong Seng Lee, and Christopher M Poskitt. 2020. Securing bring-your-own-device (BYOD) programming exams. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 880–886.

[34] Bastian Kuppers, Florian Kerber, Ulrike Meyer, and Ulrik Schroeder. 2017. Beyond lockdown: towards reliable e-assessment. *Bildungsräume 2017* (2017).

[35] Kelly Linden and Prue Gonzalez. 2021. Zoom invigilated exams: A protocol for rapid adoption to remote examinations. *British Journal of Educational Technology* 52, 4 (2021), 1323–1337.

[36] G Masilo, Sibongile Simelane-Mnisi, Andile Mji, and Ingrid Mokgobu. 2021. Students' behavioural intention and challenges to bring your own device (BYOD) in higher education during COVID-19 and beyond. *World Trans. Eng. Technol. Educ* 19 (2021), 10–15.

[37] Donald Moncrief and Ramsey Foster. 2012. Well That Was Easy: Misdirecting Respondus LockDown Browser for Fun and Profit. (2012).

[38] Robyn Muldoon. 2012. Is it time to ditch the traditional university exam? *Higher Education Research & Development* 31, 2 (2012), 263–265.

[39] Marino Nader, Ronald F DeMara, Adrian Tatulian, and Baiyun Chen. 2019. Quantitative impact on learning achievement by engaging high integrity testing using lockdown assessment for online delivery. In *ASEE Southeast Section Conferences*.

[40] Cynthia G Parshall, Judith A Spray, John Kalohn, and Tim Davey. 2002. *Practical considerations in computer-based testing*. Springer Science & Business Media.

[41] Pablo Saa, Oswaldo Moscoso-Zea, and Sergio Lujan-Mora. 2017. Bring your own device (BYOD): Students perception — Privacy issues: A new trend in education?. In *2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)*. 1–5. https://doi.org/10.1109/ITHET.2017.8067824

[42] Ammar H Safar. 2018. BYOD in Higher Education: A Case Study of Kuwait University. *Journal of Educators Online* 15, 2 (2018), n2.

[43] Vadim A Shakhnov, Lyudmila A Zinchenko, Elena V Rezchikova, and Andrei A Glushko. 2015. An opportunity in engineering education: Russian BYOD tendencies: BMSTU case study. In *2015 International Conference on Interactive Collaborative Learning (ICL)*. IEEE, 299–304.

[44] Mariana Silva, Matthew West, and Craig Zilles. 2020. Measuring the score advantage on asynchronous exams in an undergraduate CS course. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 873–879.

[45] Thea Marie Søgaard. 2016. *Mitigation of cheating threats in digital BYOD exams*. Master's thesis. NTNU.

[46] Diana Steger, Ulrich Schroeders, and Timo Gnambs. 2018. A meta-analysis of test scores in proctored and unproctored ability assessments. *European Journal of Psychological Assessment* (2018).

[47] Janus van As and Kanita Elizabeth Brits. 2022. "I get higher marks via BYOD" -a descriptive qualitative study on student experiences using BYOD e-assessments to enhance the dimensions of administration and support for learning excellence. (2022).

[48] Katharine E Welsh, Alice L Mauchline, Derek France, Victoria Powell, W Brian Whalley, and Julian Park. 2018. Would Bring Your Own Device (BYOD) be welcomed by undergraduate students to support their learning during fieldwork? *Journal of Geography in Higher Education* 42, 3 (2018), 356–371.

[49] Laura Zavala and Benito Mendoza. 2018. On the use of semantic-based aig to automatically generate programming exercises. In *Proceedings of the 49th ACM technical symposium on computer science education*. 14–19.

[50] Craig Zilles, Robert Timothy Deloatch, Jacob Bailey, Bhuwan B Khattar, Wade Fagen-Ulmschneider, Cinda Heeren, David Mussulman, and Matthew West. 2015. Computerized testing: A vision and initial experiences. In *2015 ASEE Annual Conference & Exposition*. 26–387.

[51] Craig B Zilles, Matthew West, Geoffrey L Herman, and Timothy Bretl. 2019. Every University Should Have a Computer-Based Testing Facility.. In *CSEDU (1)*. 414–420.