Transforming Computer-Based Exams with BYOD: An Empirical Study

Maximilian Sölch
maximilian.soelch@tum.de
Technical University of Munich
Munich, Germany

Markus Paulsen
markus.paulsen@tum.de
Technical University of Munich
Munich, Germany

Stephan Krusche
krusche@tum.de
Technical University of Munich
Munich, Germany

ABSTRACT

The assessment landscape in computing education is at a pivotal juncture, necessitating research-backed innovations to align with evolving pedagogical requirements. Paper-based exams, despite their historical prominence, present substantial limitations in assessing coding skills and knowledge. This paper rigorously interrogates these limitations and advocates for a paradigm shift towards computer-based assessments, with a particular focus on Bring Your Own Device (BYOD) configurations. It shows an implementation of a special exam mode in the Artemis learning platform, specifically designed for BYOD contexts with a process that is scalable to exams with more than thousand students.

In a field study employing action research, we used Artemis in three separate large-scale exams, involving 920 students in total, and gathered both quantitative and qualitative data through an online survey. The empirical evaluation of the data reveals a marked preference among students for computer-based assessments in computing education, specifically when utilizing personal setups. Findings indicate a reduced possibility in instances of academic dishonesty as compared to remote exam environments. This research substantiates the potential of computer-based exams in BYOD scenarios to rectify the incongruence between learning activities and assessment methods, thereby making a significant contribution to enhancing learning outcomes in computing education following constructive alignment.

CCS CONCEPTS

• **Applied computing → Education**; • **Social and professional topics → Student assessment**.

KEYWORDS

Computing Education, Computer-Based Assessment, Academic Integrity, Constructive Alignment, Large-Scale Examinations, Educational Technology

ACM Reference Format:
Maximilian Sölch, Markus Paulsen, and Stephan Krusche. 2024. Transforming Computer-Based Exams with BYOD: An Empirical Study. In *24th Koli Calling International Conference on Computing Education Research (Koli Calling '24), November 12–17, 2024, Koli, Finland*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3699538.3699560

Koli Calling '24, November 12–17, 2024, Koli, Finland

ACM ISBN 979-8-4007-1038-4/24/11.
https://doi.org/10.1145/3699538.3699560

1 INTRODUCTION

The evolution of computing education is pivotal in shaping the future of technology, and the assessment methods employed play a crucial role in this educational process [14]. Traditional paper-based exams have long been the standard method of assessment; however, their efficacy in computing education is increasingly being questioned [3, 31].

Paper-based exams in computing education present a paradox: While students predominantly engage in learning activities on computers, their knowledge and skills are still often assessed on paper. This misalignment between learning activities and assessment methods is exacerbated by the fact that programming on paper significantly differs from programming on a computer with an integrated development environment (IDE), as the latter brings up its own set of challenges and advantages [8, 34]. Constructive alignment [7], a principle that posits that learning activities and assessment tasks should be aligned with the learning outcomes, is thus compromised.

The COVID-19 pandemic acted as a catalyst, forcing universities to adapt to computer-based remote exams. However, post-pandemic, there has been a regression to paper-based exams, primarily due to concerns over increased cheating possibilities in unsupervised or remotely supervised environments [12, 29]. This reversion overlooks the significant weaknesses and limitations of paper-based exams in the context of computing education [15, 22]. There is a pressing need to scientifically motivate and explore alternative assessment methods that are not only secure but also pedagogically sound, aligning well with the learning activities and objectives of computing education.

This paper describes the inherent limitations of paper-based exams in computing education and their impact on constructive alignment. It explores the potential of computer-based exams, particularly in BYOD scenarios, to achieve better alignment between learning activities and assessment. We present an exam mode in the learning and research platform Artemis [19], designed for use in BYOD scenarios. We share insights from three different BYOD exams, each conducted in multiple large lecture halls with up to 450 students at the Technical University of Munich (TUM).

The main contributions are empirical evidence of the feasibility and effectiveness of computer-based exams in BYOD scenarios in large-scale computing education environments. The evaluation offers original insights into student preferences and the practical implications of implementing BYOD exams. The primary findings indicate a significant preference among students for computer-based exams in the context of programming exercises, particularly when utilizing their own familiar computing setups. The data suggests that the incidence of academic dishonesty is reduced in these settings compared to remote exam environments. The study reveals the potential of computer-based exams in BYOD scenarios to address the misalignment between learning activities and assessment methods, thereby enhancing learning experiences and outcomes in computing education.

This paper is organized to provide a comprehensive understanding of the research. Section 2 shows an overview of related work. Section 3 details the concept of BYOD computer-based exams. Section 4 delves into the methodology employed in the study. It presents the findings from implementing BYOD exams and shares insights gathered from a student questionnaire and discuss the implications of these findings on computing education. Section 5 concludes with recommendations for educators and institutions seeking to implement computer-based exams in computing education.

2 RELATED WORK

Prior research has already delved into evaluating computer-based exams, encompassing BYOD assessments, compared to conventional paper-based exams [10, 11].

Califf and Goodwin [10] introduced a laboratory final exam into a CS1 course presenting initial evidence for the effectiveness of computer-based exams in programming courses, and describe encountered problems and their solution. Rajala et al. [31] integrated the automatic assessment of electronic exams into two programming courses, specifically an introductory Java programming course and an advanced object-oriented programming course. The outcomes of their evaluation, coupled with student feedback, support the efficacy of electronic exams as a valuable tool for evaluating students in programming courses, and they recommend the adoption to other educators. Similar, Zilles et al. [35] explored the impact of a Computer-Based Testing Facility at the University of Illinois, finding that it significantly improved the assessment process for both students and instructors, and that computer science students preferred computer-based exams over paper-based exams, which aligns with the findings of this study.

Hammer et al. [17] introduce the rationale and objectives behind adopting computer-based laboratory exams. They applied the lab exam concept in one of their introductory computer science courses and presented findings from a questionnaire-based study that examined students' perceptions of this novel exam format. They conclude that computer-based exams are valuable for assessing pertinent computer programming competencies. Unlike the evaluation performed in this study, the four previously discussed publications let their students utilize university-provided computers within laboratory settings - instead of BYOD - and exclusively involved undergraduate participants in their surveys.

Nardi et al. [28] investigate the challenges of evaluating large university classes by comparing computer-based testing, specifically BYOD tests, with traditional paper-based testing. Based on a study conducted at the University of Florence in a course unrelated to programming education, the research finds that computer-based tests, facilitated by students' own devices, yielded better performance and positively correlated with perceived self-efficacy. Additionally, students expressed satisfaction with the electronic system, particularly appreciating immediate feedback.

Küppers explored the growing digitization of teaching methods and noted a lag in adopting electronic assessments [23–25]. His study attributes this gap to concerns over fairness, reliability, and costs. It suggests BYOD as a solution, particularly for financial constraints, and aims to develop a secure and reliable framework for electronic exams on students' devices. The work in this paper builds on this idea by empirically evaluating the effectiveness of BYOD in large-scale assessments, with a focus on pedagogical alignment and student preferences.

Many learning management systems incorporating an integrated exam mode alongside specialized computer-based exam tools exist [18, 21]. The exam software EXaHM is one example of a specialized computer-based exam tool [9]. This software guarantees the absence of network connectivity on laboratory computers during exams and restricts students to utilizing only the tools explicitly authorized by the instructors. Inhibiting the operation of programs such as Windows Explorer and the command window significantly diminishes opportunities for academic dishonesty. EXaHM was employed in the aforementioned evaluation conducted by Hammer et al. (2018) [17], and during the COVID-19 pandemic, its utility was subsequently extended to also facilitate remote exams [20]. Another example of an examination environment is the open-source, cross-platform software BlueBook, a Java-based application running on the student's computer [30]. It prevents students from accessing other applications or the internet during the exam and includes a crash recovery mode in case the application crashes.

Allowing students to bring their own device for writing an exam poses cheating-related issues [13, 33]. There are several approaches to prevent cheating during computer-based exams. Kurniawan et al. [22] introduce the concept of lockdown browsers for BYOD exams. Installing this special-purpose browser on the students' devices turns it into secure workstations with limited system or internet access for the period of the exam. An example instance for such a lockdown browser is the Safe Exam Browser (https://safeexambrowser.org). Another approach makes use of automated video proctoring [27]. The students' screen video recordings are compared for visual similarity of successive frames to detect changes in screen content, which would indicate that the student left the designated exam window.

Chirumamilla et al. [12] investigate the transition from traditional pen-and-paper exams to computer-based exams and assess how teachers and students perceive the ease of cheating in different exam formats. They reveal that both groups find cheating to be more accessible in computer-based exams, particularly in BYOD exams, while also acknowledging that some cheating countermeasures may be more practical in electronic exam settings. In contrast, this work proposes using exercise variants and onsite human monitoring to mitigate cheating and create a familiar development environment in computing education. Utilizing this approach avoids the need for intrusive proctoring tools, generating significant privacy concerns and false positives. This also aligns with the recent recommendations by Gulati et al. [16], who compared the security of three different proctoring regimens for BYOD exams.

3 BYOD CONCEPT

Paper-based exams fall short in facilitating competency-oriented assessments, particularly in the realm of programming education. An alternative solution would involve utilizing university-owned laptops, allowing for detailed tracking of student activities and imposing restrictions on access to resources and the internet. However, the feasibility of this approach is compromised due to its resource- and cost-intensive nature, encompassing challenges related to the acquisition and maintenance of a large number of devices.

The implementation of the BYOD concept allows students to work on the exam within the lecture hall using their personal device with their own setup as opposed to traditional pen-and-paper or provided device methods. This approach yields several advantages: Students are able to work with their personal machines, so they benefit from a familiar environment, tailored to their needs, with which they are proficient and efficient in operating. When faced with programming exercises, they are afforded the flexibility to utilize their preferred development environment, including their favorite IDE. They can harness customized keyboard shortcuts or locally stored code snippets to accelerate their programming tasks. Given that students are granted permission to access any non-intelligent resources, excluding artificial intelligence and collaborative sources, available on their computer or via the Internet, it follows that the BYOD exam is inherently designed as open-book or "open-internet" assessments. This affords students the opportunity of querying their preferred non-intelligent information sources to solve the exam. At the same time, universities are relieved of space resources, as they now only have to provide a smaller number of devices themselves [36].

However, the successful execution of the BYOD exam concept necessitates the provision of a suitable infrastructure and setup. As BYOD exams take place on the university premises, the lecture hall or exam room must be outfitted with a sufficient number of robust and reliable power outlets to obviate concerns of laptops running out of battery during the exam. The availability of a robust and reliable internet connection, whether through wireless (Wi-Fi) or wired (Ethernet) means, is imperative to ensure seamless operation. Having all internet access points within a specific IP range makes it easier to identify illegal access from an unmonitored area.

In addition to the prerequisites for the exam rooms, various requirements also exist for the exam platforms (primarily LMSs), where the actual exam exercise processing takes place. The platform must be scalable to handle more than a thousand users simultaneously. It must be resource-efficient to also allow seamless exam conduction for students with weaker devices to not disadvantage them. It should monitor exam metadata for each student, including IP addresses, browser fingerprints, and session IDs to identify attempts of academic dishonesty.

The open-source platform Artemis[1] fulfills all of these requirements [19]. It runs in the browser and supports various exercise types, including programming, modeling, text, generic file upload exercises and quizzes. Programming exercises and quizzes can be graded automatically, while semi-automatic grading methods using machine learning are available for modeling and text exercises. For programming exercises, students push code to individual Git repositories. Students submit other exercise types directly in the browser. The exam mode in Artemis supports the same exercise types and has demonstrated its worth for computer-based exams with up to 1,500 students [26]. Instructors used Artemis for more than 20 supervised BYOD exams on the university's premises, all of which went smoothly. The evaluation in Section 4 focuses specifically on three exams, in which the instructors followed the BYOD exam process shown in Figure 1.

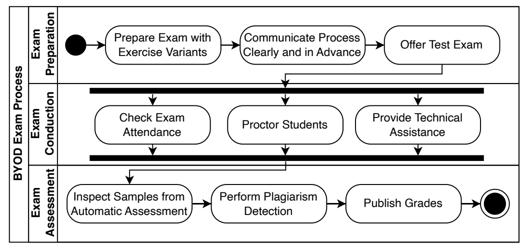

Figure 1: UML Activity Diagram depicting the BYOD Exam Process, encompassing stages of preparation, conduction, and assessment

3.1 Exam Preparation

The exam mode allows instructors to define exercise variants, as shown in Figure 2, so that the students do not get the exact same version of the exam, but some variations that Artemis assigns automatically to each individual student. This approach serves to diminish the potential for cheating, as students are unable to merely read and replicate content from the screens of their neighboring peers.

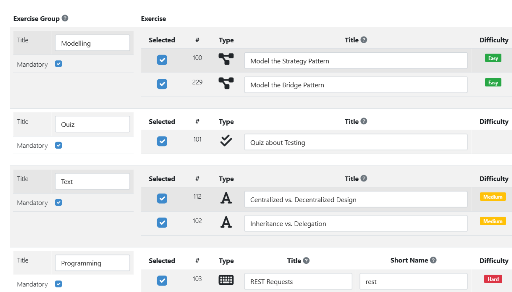

Figure 2: Instructor's configuration view for multiple exercise variants in Artemis

The exam mode seamlessly integrates with Artemis, a platform familiar to students through its use in the course as a learning platform. This familiarity allows them to effortlessly navigate and access course materials, including lecture slides and exercises, during the exam, eliminating the need to visit an external website. Supervisors can employ Artemis' announcement feature to promptly and synchronously communicate important information to exam participants in the event of unforeseen circumstances that may impact the examination process. Artemis offers extensive analysis capabilities to aid in detecting academic dishonesty, including the ability to identify suspicious activities on the platform, e.g. detecting if someone participates in the exam from outside the supervised exam rooms.

[1] Artemis: Interactive Learning with Individual Feedback - https://github.com/ls1intum/Artemis

3.2 Exam Conduction

Students may only use one device with a single set of input and output devices (e.g. only the use of a screen, mouse and keyboard is permitted; all other devices, including smartphones, must be switched off). During the exam, supervisors offer students technical support, but are not allowed to answer questions about the exam content. To discourage academic dishonesty, supervisors move around the exam room and carefully inspect the content displayed on students' screens to detect signs of unauthorized aids or communication. They should not stand too close to any single student and move quietly to avoid creating a distress situation for affected students. In the case of clearly suspicious behavior, they secure photographic evidence and then inform the student about the misconduct and the expected consequences.

Supervisors are responsible for executing an attendance verification process. This involves checking that each student is sitting in their assigned seat, is logged into the correct account in Artemis and has a valid student ID or driver's license. Students must confirm their attendance with a signature. To speed up and simplify this testing process for both supervisors and students, we have developed an iPad application that works with Artemis.

Figure 3: Exam attendance app supporting supervisors to perform the necessary attendance verification process

Figure 3 shows the view presented to supervisors, incorporating pertinent student name, picture, and university identifier. This data must align with the physical identification presented by the student and with their active session in their web browser when using Artemis. For additional fraud protection, it is recommended to let the student open the user settings on Artemis before checking the university identifier on the platform to render a potential manipulation of the displayed HTML useless. In case all elements correspond accurately, students can validate their attendance for the exam by providing their signature on the iPad using an Apple Pencil.

There are two ways of addressing technical problems in the exam: prevention and mitigation. In the course of exam preparation and execution, it makes sense to make use of both options simultaneously. Prevention is achieved by informing the students about the upcoming setting with the explicit request to check all important functionalities (e.g. correctly working IDE) once more the day before. To familiarize students with the overall exam setting, it is crucial to conduct a test exam prior to the actual exam. Mitigation is a two-stage process that is utilized in the event of technical problems occurring on a student's machine during the exam. Stage one consists of the supervisory staff attempting to provide simple support, not on a quality-of-service but on a best-effort basis, taking into account supervisory capacity. This means that if there is no quick solution in sight, stage two comes into effect, which consists of the affected student going to a separate supervised room and trying to continue the exam there on university computers. Artemis continuously saves the current progress so that students can continue on a second device without loosing data. Supervisors can easily grant individual time extensions to compensate the lost time during the technical issue. Students are advised to try out the setup on the university computers before the exam starts. However, the general principle holds: students are responsible for their own technical problems, unless they can prove otherwise.

3.3 Exam Assessment

Following the exam conduction, Artemis facilitates (semi-) automatic assessment capabilities that accelerate the correction process [4, 5]. This also includes a procedure for reviewing exam results in which students can object to certain assessments if they consider them to be incorrect. Instructors can review samples from the automatic assessment to ensure assessment accuracy. Artemis seamlessly conducts plagiarism detection across all submissions [6]. Instructors can filter submissions with high similarity and determine the appropriate course of action. The final step is to publish the grades within Artemis or export them for integration into the university's grade management system.

4 EVALUATION

In this section, we outline the methodology employed to evaluate the viability of the proposed BYOD concept following a mixed method approach. The concept was introduced and implemented in the final exams of three distinct computer science courses at TUM during the winter semester 2022-23 as part of a field study:

1. *Patterns in Software Engineering (PSE)*: 100 min exam with 453 participants in the final.
2. *Introduction to Programming (CS1-M)*: 100 min exam for management students with 418 participants in the final.
3. *Introduction to Informatics (CS1-C)*: 100 min exam for computer science students with 49 participants in the final.

In total, 920 students participated in the three finals. Each exam encompassed quizzes alongside interactive modeling and programming exercises. For the exams, students are permitted to utilize external resources; nonetheless, it is imperative that they independently address and resolve the exam materials. Consequently, any form of communication or collaboration with other exam participants or external parties is strictly prohibited, and the utilization of AI tools, including but not limited to ChatGPT[2] and GitHub Copilot[3], is expressly disallowed. The creation of exercises involved careful consideration to design them in a way that discourages facile resolution with the aid of artificial intelligence. In the following, we describe the research objectives, the study setup, the results, and the findings of the evaluation which is based on action research [2]. We then discuss the implications of the findings and the limitations of the evaluation.

[2]https://openai.com/blog/chatgpt
[3]https://github.com/features/copilot

4.1 Research Objectives

The evaluation aims to address three research questions to deepen the understanding of assessments in computing education, particularly in the context of (BYOD) exams:

RQ1 What are the main (technical) problems that students have to deal with when writing a BYOD exam?
RQ2 What are the preferences of students when it comes to paper-based and computer-based exams and to remote or onsite supervision?
RQ3 What kind of cheating possibilities do the different scenarios exhibit?

RQ1 explores the main technical challenges students face during a BYOD exam, as identifying these issues is crucial for the seamless implementation of technology-driven assessments. RQ2 delves into student preferences concerning paper-based versus computer-based exams and the type of supervision (remote or onsite). Understanding these preferences can offer insights into how to design assessments that are both effective and accepted by students. RQ3 examines the potential avenues for cheating in different assessment scenarios. This investigation is vital for ensuring the academic integrity of the exam process, which in turn influences the reliability and validity of the assessment outcomes. Based on these research questions, we state the following four hypotheses together with the relevant survey questions:

- **H1 Preference**: Students prefer computer-based exams (supervised, onsite) over paper-based exams
  - **P-Q1** Overall, which exam scenario would you prefer for exams which include programming exercises?
  - **P-Q2** Overall, which exam scenario would you prefer for exams which do not include programming exercises?
  - **P-Q3** Would you like to write more exams as computer-based exams in the lecture hall?
- **H2 Cheating**: Students attempt to cheat less likely in supervised exams onsite than in remote exams
  - **C-Q1** Cheating is easier on a computer-based exam in the lecture hall than on a paper exam.
  - **C-Q2** Cheating is easier on a computer-based exam at home than in a supervised exam in the lecture hall.
  - **C-Q3** I thought about possible cheating options before taking the computer-based exam in the lecture hall.
- **H3 Issues**: With a dedicated infrastructure and tool support, students do **not** experience relevant issues
  - **T-Q1** How was your general experience with the computer-based exam on Artemis?
  - **T-Q2** Have there been any technical problems?
- **H4 Setup**: Students prefer their personal and familiar setup in a computer-based exam
  - **S-Q1** I prefer to use my own computer rather than a computer provided by the university.
  - **S-Q2** I prefer to use my own computer because I am more familiar with its functions and operation.

H1 asserts that students would favor supervised, onsite computer-based exams over traditional paper-based assessments. This hypothesis aims to explore how technological advancements in educational settings align with student preferences. H2 posits that instances of academic dishonesty are less likely in supervised, onsite exams compared to remote settings, addressing concerns around the integrity of assessments. H3 anticipates that with a dedicated infrastructure and proper tool support, students will not encounter significant technical disruptions, thereby testing the operational feasibility of BYOD exams. H4 suggests that students have a preference for using their personal and familiar computing setups during computer-based exams, offering insights into comfort and familiarity factors that may impact performance. Each of these hypotheses serves to dissect critical elements that contribute to the effectiveness, reliability, and acceptability of computer-based assessments in computing education.

4.2 Study Setup

To gain comprehensive insights into the experiences and perspectives of the students regarding the newly introduced BYOD assessment concept, we employed a survey-based approach. All 920 students who participated in the aforementioned final exams were invited via e-mail to take part in the survey. This anonymous survey, which was conducted in March 2023 and hosted on the community version of the open-source survey tool LimeSurvey[4], aimed to collect valuable data pertaining to their experiences and perceptions of the BYOD assessment concept. We sent three reminders to the students who had not yet participated. All survey questions, except for the introductory demographic queries and certain open-ended text responses, employ a 5-point Likert scale [1]. The survey commences with a series of introductory demographic questions, encompassing identification of the specific exam taken among the three options, the student's current study program, ongoing degree pursued, and academic semester. Additionally, respondents are asked to provide information regarding the hardware employed during the exam, including details on the operating system and web browser used.

Following the demographic section, the survey includes questions related to the participants' experience with computer-based exams. This section examines their familiarity with computer-based exams, both within the university setting and in remote environments at home. Subsequently, the survey addresses participants' specific experiences with computer-based exams conducted through Artemis. It seeks to gauge their impressions and observations concerning the usage of this platform for assessment.

The survey proceeds to investigate aspects related to the editing of the exam content and the participants' perceptions of the potential for cheating in the context of computer-based exams. Participants are then queried about their preferred location for taking exams, shedding light on whether they prefer on-campus exam halls or remote settings. The subsequent section pertains to the technical equipment available to participants and the adequacy of infrastructure within the lecture hall during the exam. Following structured questions, the survey includes open-ended queries, inviting participants to articulate their perspectives on both the advantages and disadvantages associated with the implemented exam format. Finally, the survey concludes with a set of questions aimed at capturing participants' overall impressions of the exam experience and their preferences for assessment formats in future

[4]https://www.limesurvey.org

educational endeavors. Of the 920 invited participants, 467 completed it (response rate 51%).

4.3 Results

In the context of the demographic questions, it is discerned that the majority of respondents are pursuing degrees in management (40%), closely followed by students majoring in computer science (40%). A smaller proportion of participants are enrolled in programs related to information systems (6%), information engineering (6%), and other miscellaneous degree programs (8%). The distribution of respondents between bachelor's (43%) and master's (57%) programs is nearly equitable. Furthermore, a predominant number of participants engaged in the *PSE* exam (51%), with a substantial minority undertaking the *CS1-M* exam (43%) and a smaller cohort participating in the *CS1-C* exam (6%). Among undergraduate students, around 12% are enrolled in their first, 19% in their second, 33% in their third, 28% in their fourth and 8% in their fifth year of study. For graduate students, around 50% are enrolled in their first, 34% in their second, 13% in their third and 3% in their fourth year of study or later.

Regarding the preferred operating systems employed on their laptops, the following distribution is observed: macOS (50%), Windows (43%), Linux (7%). Similar for web browsers: Google Chrome (56%), Safari (20%), Firefox (15%), Microsoft Edge (5%), and others (4%). The last question in this section solicited participants' self-assessment of their programming proficiency. The majority of respondents characterized their programming experience as intermediate (34%), with a notable proportion rating themselves as advanced (31%). A smaller fraction of participants identified as beginners (25%), experts (5%), or novices (5%) in the realm of programming proficiency.

4.3.1 General Experience. The survey asked about the participants' familiarity with computer-based exams. Most respondents (85%) reported previous experience with computer-based exams conducted within the lecture hall using their devices. In contrast, only a few (12%) indicated having encountered computer-based exams exclusively within computer laboratories. Furthermore, participants were questioned regarding their frequency of engagement in computer-based exams conducted remotely at home. A notable segment reported involvement in either a substantial number exceeding five exams (27%) or, conversely, no previous experience (27%) with remote computer-based exams. The remaining respondents (46%) fell within the spectrum of having participated in at least one but fewer than six such exams from their home.

4.3.2 Experience in Computer-based Exam on Artemis. Students' overall experience with the computer-based exam on Artemis (**T-Q1**), as depicted in Figure 4, has received positive feedback. A significant majority of participants characterized their general experience with the computer-based exam on Artemis as either "very good" (49%) or "good" (39%). In contrast, a relatively small percentage of participants found the experience to be "moderate" (10%), "poor" (2%), or "very poor" (0%). Most participants (88%) reported an absence of technical problems during their exam (**T-Q2**), as illustrated in Figure 5. Only a minority disclosed experiencing challenges associated with their personal hardware (5%), software (4%), network connectivity (3%), or other miscellaneous issues (3%). Some of the issues in the respective free text answer field mentioned submission problems shortly before the deadline, problems with the building system or long loading times. Participants voiced a consensus on the benefits of undertaking computer-based exams using their computing setups, with a majority expressing strong agreement (56%) and agreement (32%). A smaller percentage adopted a neutral stance (9%), while a minority expressed disagreement (2%) or strong disagreement (1%). Moreover, the study participants attested to the intuitiveness and user-friendliness of the exam mode. A substantial proportion strongly agreed (42%) and agreed (45%) with this statement, with a modest number maintaining a neutral stance (11%), while a small minority disagreed (1%) or strongly disagreed (1%).

When asked about the advantages of solving programming exercises during the exam using an IDE, most participants endorsed this approach, with strong agreement (61%) and agreement (29%). The remainder conveyed neutrality (8%), disagreement (1%), or strong disagreement (1%). Additionally, students conveyed a sense of increased confidence when accessing pertinent content such as definitions, lecture materials, example codes, and related resources during the exam. The responses indicate strong agreement (60%), agreement (27%), neutrality (10%), disagreement (2%), and strong disagreement (1%) on this matter.

The responses varied when asked to assess whether computer-based exams are more challenging than traditional paper-based exams. While a minority strongly agreed (6%) or agreed (10%) with this assertion, a substantial segment adopted a neutral stance (38%). Conversely, a notable percentage disagreed (33%), with an additional fraction expressing strong disagreement (13%).

4.3.3 Editing and Cheating. The survey asked participants for their perspectives on the potential for academic misconduct. Figure 6 summarize the results. Regarding the ease of cheating on computer-based exams in the lecture hall as opposed to traditional paper-based exams (**C-Q1**), the responses varied. While a fraction held strong agreement (3%) or agreement (14%) with this notion, a substantial number adopted a neutral stance (31%). In contrast, a notable proportion disagreed (33%), with an additional segment expressing strong disagreement (19%). Similarly, when participants were queried about the perceived disparity in the susceptibility to cheating between computer-based and paper-based exams (**C-Q2**), responses showed diversity. A significant contingent demonstrated strong agreement (23%) or agreement (38%), while a minority maintained neutrality (20%). Conversely, a portion exhibited dissent, with 12% in disagreement and 7% strongly disagreeing with the premise. Regarding contemplating potential cheating methods before undertaking computer-based exams in the lecture hall (**C-Q3**), the majority expressed strong disagreement (54%) or disagreement (29%). A smaller proportion held a neutral stance (9%), while a minority concurred with this notion, with 7% in agreement and 1% strongly agreeing. Participants were asked whether they believed consistent and direct supervision, including instructors circulating through the lecture hall, made cheating difficult. A majority affirmed this statement, with 24% in strong agreement and 42% in agreement. A fraction remained neutral (24%), 8% disagreed, and 2% strongly disagreed. The final question in this group about cheating investigated whether participants thought cheating was more feasible during computer-based exams than paper-based counterparts.

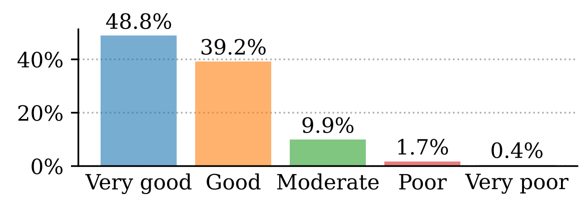

Figure 4: How was your general experience with the computer-based exam on Artemis? (T-Q1)

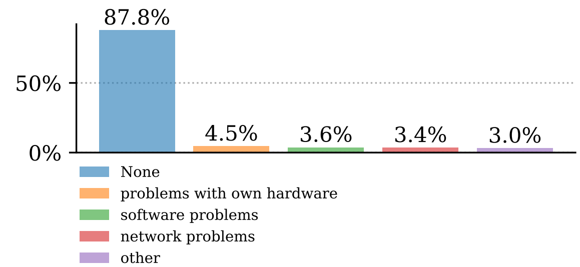

Figure 5: Experienced issues with computer-based exams on Artemis (T-Q2)

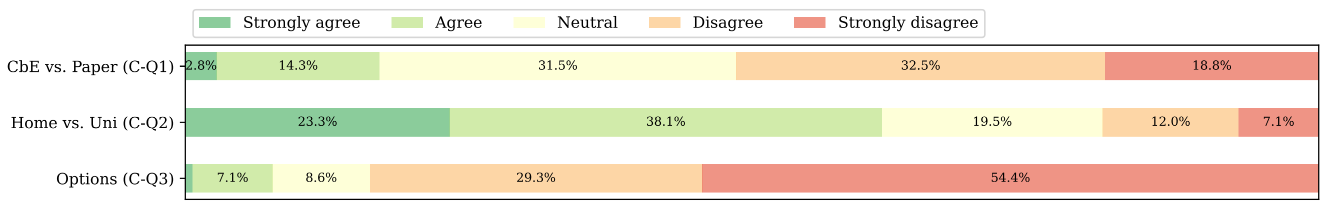

Figure 6: Perceptions of cheating viability in computer-based exams (C-Q1, C-Q2, C-Q3)

The responses encompassed strong agreement (6%), agreement (29%), neutrality (32%), disagreement (22%), and strong disagreement (11%).

4.3.4 Organization. The survey continued with questions about students' perceptions regarding the practicality of computer-based exams conducted at home versus in the lecture hall. The participants expressed diverse views in response to whether they believed it would be more convenient to take a computer-based exam at home compared to one in the lecture hall. Notably, 23% strongly agreed, while 30% agreed, with a sizable segment remaining neutral (27%). Conversely, 16% disagreed, and 4% strongly disagreed with this assertion. This question group investigated their comfort levels within the lecture hall environment, particularly when taking an exam in the presence of their peers. A portion displayed strong agreement (13%), 26% agreed, and 31% stayed neutral. The remaining participants, however, exhibited disagreement (21%), with 9% strongly disagreeing with the statement. A subsequent query concerned participants' ability to concentrate when alone at home, as opposed to within a collective exam setting. Responses indicated that 22% strongly agreed, while an equivalent proportion agreed they could concentrate better when isolated at home. A contingent of 27% maintained a neutral standpoint, while 22% disagreed and 7% strongly disagreed.

4.3.5 Technical Equipment. In the following questions, the survey investigated students' perspectives on their technical equipment and the university's infrastructure. When queried about their preference for using their personal computer over a university-provided computer during computer-based exams in the lecture hall (**S-Q1**), the vast majority indicated agreement, with 71% strongly agreeing and 22% agreeing. A small fraction remained neutral (6%), while a minority expressed disagreement (1%) or strong disagreement (0%). Additionally, participants emphasized their preference for using their own computers for computer-based exams due to their familiarity with their own device's functions and operation (**S-Q2**). The majority strongly agreed (76%) or agreed (19%), with a minimal percentage maintaining neutrality (5%). The survey also probed participants if they are concerned when using a computer (own or provided) during a computer-based exam that technical problems will occur that they cannot immediately fix. Responses revealed that 20% strongly agreed with this message, 43% agreed, while 17% maintained neutrality. Conversely, 16% disagreed, and 4% strongly disagreed with this concern. Participants generally perceived the lecture halls as suitable for computer-based exams, with 19% strongly agreeing and 45% agreeing. A proportion assumed a neutral stance (21%), while 12% disagreed, and 3% strongly disagreed with this statement. Regarding the adequacy of their personal computers for participating in computer-based exams, a majority expressed strong agreement (55%) or agreement (37%). A small percentage adopted a neutral perspective (5%), while 2% disagreed, and 1% strongly disagreed. For the last question in this category, participants were asked to assess the reliability of the internet connection in the lecture room. Responses indicated a strong consensus, with 47% strongly agreeing and 46% agreeing. A minor percentage remained neutral (6%), only 1% disagreed, and nobody strongly disagreed.

4.3.6 Advantages and Disadvantages. When asked about the advantages of computer-based exams in the lecture hall, participants mainly mentioned the ability to code within an IDE, perceived as more advantageous and closer to reality than coding on paper.

> *"A true open-book exam is closer to reality, based on an actual understanding of the subject instead of remembering stuff anyone anywhere else would just google IRL. Especially for (semi) applied subjects it's like day and night in how much better computer-based exams are." - Participant*

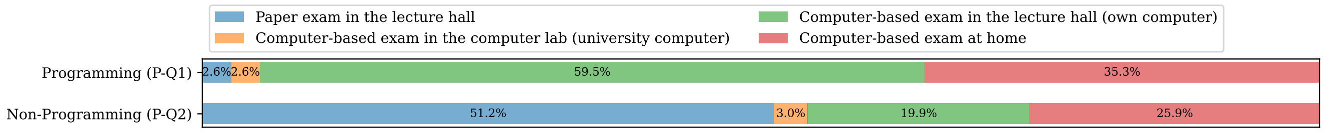

Figure 7: Preferences for exams with and without programming exercises (P-Q1, P-Q2)

The availability of supervisors for technical support during the exam was noted as a benefit. Participants also emphasized that writing the exam collaboratively in the lecture hall provides a sense of "not being alone," which increases awareness of the exam situation. Computer-based exams in the lecture hall were seen as reducing the likelihood of cheating and providing a sense of security against unwarranted plagiarism accusations due to the presence of invigilation and surveillance.

> *"I see it as the best of both worlds. Computer-based exams in programming focused subjects are generally much closer to the real life challenges of programming. Paper exams and programming on paper do not resemble the real world challenges very well. Writing the exam in the lecture hall reduces the ability to cheat to an acceptable level. While depending on the kind of supervision, exams at home often made it very easy to cheat, to a point where it felt like a disadvantage to take the exam without any cheating, because it felt reasonable to assume, that a lot of people would take the opportunity." - Participant*

Only a few study participants mentioned concerns associated with computer-based exams in the lecture hall: These concerns include, above all, the increased susceptibility to technical difficulties that could disrupt the examination process. Issues related to internet connectivity and the availability of charging facilities were identified as potential drawbacks. Participants also expressed reservations about the suitability of the desks within certain lecture halls at TUM, citing concerns about surface size and tilt, which could impact their ability to effectively use a computer during the exam. Supervisors moving throughout the lecture hall and the auditory distractions generated by typing on keyboards were noted as potential factors that could impede concentration during the exam.

4.3.7 Overall Impression. In the final part of the survey, participants were asked about their overall impression and their preferences for future exams. The results for their preference regarding exams that include programming exercises and exams without programming exercises are illustrated in Figure 7. Regarding exams that encompass programming exercises (**P-Q1**), a majority (60%) favored the notion of a computer-based exam in the lecture hall utilizing their personal computers, while 35% preferred remote exams at home. The remainder of the respondents indicated a preference for traditional paper-based exams or computer-based exams within a dedicated computer lab using university-owned machines. In contrast, among responses regarding preferred exam formats for assessments excluding programming exercises (**P-Q2**), most (51%) picked paper-based exams within the lecture hall, followed by 26% favoring computer-based exams at home. An additional 20% chose computer-based exams within the lecture hall utilizing their personal computers. In comparison, 3% preferred computer-based exams within a dedicated computer lab using university-owned machines. The survey's last question, as shown in Figure 8, investigated whether participants desired to write more computer-based exams within the lecture hall (**P-Q3**). 69% of respondents affirmed this inclination, while only 10% expressed dissent. The remaining 21% conveyed uncertainty on this matter.

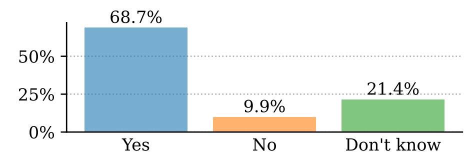

Figure 8: Would you like to write more exams as computer-based exams in the lecture hall? (P-Q3)

4.4 Findings

The investigation into using BYOD computer-based exams in programming education has yielded valuable insights that shed light on students' perceptions and experiences. Given the research objectives and hypotheses, we can now evaluate the results and assess whether they confirm the initial expectations.

At the beginning of the evaluation process, we conducted several preparatory steps to ensure the meaningfulness of the given data. We cleaned the data by removing unnecessary values to streamline the dataset for further analysis. We categorized the different study programs based on the program's content and objectives into two primary groups: "Informatics-based" and "Management-based". We reorganized the semesters into study years to better group the students into steps of their academic progress. In an effort to better understand the impact of student's exposure to computer-based exams, we categorized the number of already written computer-based exams into three distinct categories: "0", "1", and "more than or equal to 2". We merged values containing the word "very" with their respective counterparts that did not include the word to show trends instead of exact values.

Finally, we conducted a χ^2-test of independence on all dataset columns, calculating p-values, Cramer v, and Cohen ω values to assess dependency strength. We then removed question pairs with statistically insignificant associations, weak associations, and those stating obvious relations. For most of the remaining question pairs, the result of the independence test showed no or weak correlations. Only the question pairs carrying the most meaningful associations are discussed below. H1 postulates that students favor onsite computer-based exams over traditional paper-based ones. The findings indicate a marked preference among students for computer-based exams within the lecture hall using their devices. A significant majority preferred this format when it involved programming exercises. This aligns with the idea that computer-based exams, which allow the utilization of IDEs, enhance the authenticity of programming assessments. However, it's noteworthy that preferences for non-programming exercises leaned more toward traditional paper-based exams. In addition, the strong tendency to take remote exams at home should not be ignored. Therefore, we can only confirm H1 for exams with programming tasks, which aligns with previous research findings [35].

> **Finding H1a:** For programming tasks, students prefer computer-based exams rather than paper-based ones.

A non-negligible amount of students prefer exam scenarios at home. Reasons in the survey answers include less effort for traveling, a safe environment without disturbance from other students, and larger space at home. Writing the exam at home, which some of them are used to due to COVID-19, allows them to save time and to stay more concentrated. However, other students state they get more easily distracted at home and therefore prefer to take the exam in the lecture hall, showing a diverse set of opinions about the optimal exam environment. Easier cheating possibilities at home might also influence the opinion about the best setting.

> **Finding H1b:** A quarter of students prefers computer-based exams at home.

H2 hypothesizes that students would be less likely to attempt cheating in supervised onsite exams compared to remote exams. The results reflect diverse perceptions regarding the ease of cheating in different exam settings, but show a general tendency that students find it easier to cheat in remote exams at home. The majority states that they did not think about possible cheating options during the computer-based exam in the lecture hall, confirming H2. Various factors, including the effectiveness of proctoring and the integrity of the exam environment, influence perceptions of cheating. While cheating is a critical concern, its exact relationship to exam settings should be investigated further to determine whether it is influenced by factors such as supervision and personal ethics.

> **Finding H2:** Students are less likely to cheat in computer-based exams that are supervised in the lecture hall compared to remote exams, but some find it easier than in paper-based exams.

H3 states that students would not encounter relevant technical issues when provided with a dedicated setup, infrastructure, and tool support. The results indicate that a significant portion did not experience technical problems during the exam. For the minority who did encounter problems, the issues were primarily related to personal hardware, software, or network connectivity and could be solved before and during the exam with the help of the supervisors in most cases. In the rare cases where technical issues persisted, the students were sent to the lab room to finish their exam on a university computer. The time for transit was tracked and added to the individual's exam time. This suggests that a well-prepared and appropriate technical environment can indeed minimize such issues. While it is challenging to eliminate all technical hindrances, the study demonstrates that dedicated setups can effectively mitigate them, thus confirming H3.

> **Finding H3:** Students do **not** experience technical disruptions during the exam in lecture halls with proper Wi-Fi and power supplies.

The χ^2-test of independence indicates, that the association between the question "Because of the information I received in advance about the computer-based exam, I was able to process it without any problems." and question **T-Q2** is significant ($\chi^2 = 42,80;\ p = 5,39 * 10^{-6}$) and medium ($Cramer\ v = 0,22;\ k = 3$). Although the data indicates a statistically significant and medium correlation, more research needs to be done to indicate a causal relationship.

H4 predicts that students would prefer their personal and familiar setup in a computer-based exam. The survey data affirm the expectation that students prefer using their own devices for exams, citing familiarity and comfort as key factors. The majority strongly agreed with this perspective, confirming H4. Only a few expressed their preference for taking the exam on a university-provided computer.

> **Finding H4:** Students overwhelmingly favor using their personal devices during exams, citing the comfort and familiarity of their own setups.

The χ^2-test of independence indicates, that the association between the question "How would you describe your programming experience?" and "Computer-based exams are more difficult than paper-based ones." is significant ($\chi^2 = 52,07;\ p = 1,63 * 10^{-8}$) and medium ($Cramer\ v = 0,24;\ k = 3$). While the data suggests a statistically significant and moderate correlation, further research in this area is required to indicate a causal relationship.

Based on the findings, it is essential for institutions to consider implementing computer-based exams that cater to the preferences and needs of students. Adequate technical support and infrastructure are crucial to ensuring a positive experience. Additionally, strategies for minimizing cheating and academic misconduct should be developed and communicated clearly to students.

In conclusion, computer-based exams, particularly those utilizing students' personal devices in the lecture hall, promise to enhance the assessment experience in programming education. While challenges exist, the benefits, including the ability to code in an IDE and receive immediate feedback, contribute to the appeal of this exam mode. However, institutions should be attentive to technical issues and potential academic misconduct, addressing them through suitable measures and support systems.

4.5 Discussion

Based on the experiences derived from the organization of multiple exams using the presented BYOD concept, coupled with the insights garnered from the survey, we strongly advocate the integration of BYOD exams as a valuable approach for programming education assessment. However, it is imperative to underscore that the successful implementation of this concept at other academic institutions necessitates a meticulous and comprehensive setup and preparation process. This endeavor is pivotal to ensure the seamless and effective execution of exams within the BYOD framework.

The claim of increased cheating activities in BYOD exams cannot be substantiated. Throughout the three examinations, there was no discernible increase in cheating incidents. Effective examination systems mitigate such concerns by incorporating features such as tailored exercise variants for each student and facilitating instructors in promptly identifying irregularities in student submissions, including instances of plagiarism or discrepancies in session details like multiple IPs or browsers. Students facing concentration challenges can request special seating, while those with specific weaknesses or disabilities can seek individual time extensions and seating accommodations through the examination board. This is supported by the Artemis Exam Process. With appropriate examination room infrastructure and availability, combined with an effective exam system, BYOD exams can be seamlessly scaled to accommodate over 1500 students simultaneously. Artemis enhances the examination process for both students and instructors, offering familiar usability for students, while providing instructors with helpful features such as (semi-) automatic assessment, plagiarism detection, exercise variants, session anomaly detection, and more.

Computer-based exams with BYOD emphasize applying knowledge at higher cognitive levels, moving beyond rote memorization or basic comprehension of course content. Consequently, instructors must exercise vigilance in crafting exams to strike an appropriate balance, ensuring they do not inadvertently become excessively challenging or lengthy due to the accessibility of support tools, the internet, and course resources. There is a latent risk of overestimating students' capacities in this context. As a safeguard, we recommend conducting a thorough review with a representative student, using tools such as Artemis' test run feature to refine the exam's design and structure.

In case of technical issues, the supervisors should assist the students in resolving the problems. If the issue cannot be resolved, students should be sent to a computer lab to finish their exam there. The time for transit is tracked and added to the individual's exam time. This ensures that students are not disadvantaged due to technical problems. However, the possibility of technical issues or a preference for paper-based exams by some students should not be reasons to dismiss the BYOD concept. It is essential to emphasize the benefits of the BYOD concept, particularly in enhancing the authenticity of programming exams and aligning them with real-world scenarios.

4.6 Limitations

In delineating the limitations inherent in the conducted study, we adhere to the categorization framework proposed by Runeson and Höst [32], addressing potential threats to internal, external, and construct validity:

Internal Validity: A limitation pertains to the survey completion rate. Notably, a response rate of 51 % could indicate that not all student perspectives are represented equally in the results. This introduces a threat to internal validity, as the results are predicated on a potentially biased sample, comprising solely those who concluded the survey.

External Validity: Threats to external validity emerge from the study's specific context. The research was exclusively conducted within a singular university where most students possessed suitable laptops for participation in computer-based exams. The conditions and technological resources available at other universities may differ, rendering the generalization of study findings less certain. Moreover, the evaluation was confined to a single course period and exclusively encompassed students within technical management and computer science programs. Consequently, the potential variation in results for less technical disciplines restricts the broader applicability of the study.

Construct Validity: Potential issues could be associated with formulating survey questions, particularly in relation to the experimenters' expectancies. Additionally, the evaluation may be susceptible to participant bias. Notably, the absence of major technical issues during the administration of the three exams may have influenced participants to provide more positive responses. The occurrence of technical problems could have yielded divergent results. Nevertheless, it is essential to mention that this field study did not encompass deliberate experimentation with induced failures. Conversely, it is plausible that students with unfavorable exam experiences due to inadequate exam preparation could have influenced the survey results, introducing another dimension of construct validity concerns.

5 CONCLUSION

This research paper has highlighted the limitations of paper-based exams in computing education and demonstrated the advantages of computer-based exams in BYOD settings through empirical data from 467 survey responses. The findings show a clear preference for computer-based assessments among students and indicate a reduced likelihood of academic dishonesty compared to remote exams. The experience with the BYOD concept and Artemis underscores the feasibility of scalable exams, limited mainly by venue capacity and supervision resources, without increasing grading workload. We will continue using the concept and with Artemis now in use at over 20 universities, including non-technical fields. Future studies can explore BYOD across diverse contexts to address the study's limitations and advance understanding.

We recommend that educators start with a pilot test of the computer-based assessment format to identify logistical and technical challenges. Training sessions for faculty and students can ease the transition, while test exams help students adapt to the new format and infrastructure. Institutions should invest in secure and robust infrastructure, including reliable wireless connectivity and power solutions, and implement rigorous supervision and session monitoring to prevent academic dishonesty. Revising exam policies to include provisions and contingency plans for computer-based exams is crucial. By adopting these recommendations, educators and institutions can achieve a more aligned, secure, and pedagogically sound assessment method, thereby significantly enhancing the quality of computing education.

ACKNOWLEDGMENTS

This paper includes sections revised with the help of ChatGPT to enhance clarity and readability. We have carefully reviewed all text produced by AI. The dataset for the online survey discussed in this paper is available on OSF[5].

REFERENCES

[1] I.E. Allen and C.A. Seaman. 2007. Likert scales and data analyses. *Quality Progress* 40 (2007), 64–65.

[2] David E Avison, Francis Lau, Michael D Myers, and Peter Axel Nielsen. 1999. Action research. *Commun. ACM* 42, 1 (1999), 94–97.

[3] 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. In *Proceedings of the 8th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE '03)*. ACM, 16–20.

[4] Jan Philip Bernius, Stephan Krusche, and Bernd Bruegge. 2021. A Machine Learning Approach for Suggesting Feedback in Textual Exercises in Large Courses. In *Proceedings of the Eighth ACM Conference on Learning @ Scale (L@S '21)*. ACM, 173–182.

[5] Jan Philip Bernius, Stephan Krusche, and Bernd Bruegge. 2022. Machine Learning Based Feedback on Textual Student Answers in Large Courses. *Computers and Education: Artificial Intelligence* 3 (2022), 100081.

[6] Jonnathan Berrezueta-Guzman, Markus Paulsen, and Stephan Krusche. 2023. Plagiarism Detection and Its Effect on the Learning Outcomes. In *2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T)*. 99–108.

[7] John Biggs. 1996. Enhancing teaching through constructive alignment. *Higher Education* 32, 3 (1996), 347–364.

[8] Hugh Burkhardt and Daniel Pead. 2003. Computer-based assessment: a platform for better tests. *Whither assessment* (2003), 133–148.

[9] Axel Böttcher, Veronika Thurner, and Daniela Zehetmeier. 2020. Alignment of Teaching and Electronic Exams and Empirical Classification of Errors for an Introductory Programming Class. In *32nd Conference on Software Engineering Education and Training*. 1–10.

[10] Mary Elaine Califf and Mary Goodwin. 2002. Testing Skills and Knowledge: Introducing a Laboratory Exam in CS1. In *Proceedings of the 33rd SIGCSE Technical Symposium on Computer Science Education (SIGCSE '02)*. ACM, 217–221.

[11] Jacabo Carrasquel. 1985. Competency Testing in Introductory Computer Science: The Mastery Examination at Carnegie-Mellon University. In *Proceedings of the Sixteenth SIGCSE Technical Symposium on Computer Science Education (SIGCSE '85)*. ACM, 240.

[12] 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 and Evaluation in Higher Education* 45, 7 (2020), 940–957.

[13] 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.

[14] Robyn Fitzharris and Simon Kent. 2020. *Adoption of Bring-Your-Own-Device Examinations and Data Analytics*. Springer, 327–348.

[15] Scott Grissom, Laurie Murphy, Renée McCauley, and Sue Fitzgerald. 2016. Paper vs. computer-based exams: A study of errors in recursive binary tree algorithms. In *Proceedings of the 47th technical symposium on computing science education*. ACM, 6–11.

[16] 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)*. ACM, 429–435.

[17] Sabine Hammer, Martin Hobelsberger, and Georg Braun. 2018. Using laboratory examination to assess computer programming competences: Questionnaire-based evaluation of the impact on students. In *Global Engineering Education Conference*. IEEE, 355–363.

[18] Petri Ihantola, Tuukka Ahoniemi, Ville Karavirta, and Otto Seppälä. 2010. Review of Recent Systems for Automatic Assessment of Programming Assignments. In *Proceedings of the 10th Conference on Computing Education Research (Koli Calling)*. ACM, 86–93.

[19] Stephan Krusche and Andreas Seitz. 2018. ArTEMiS: An Automatic Assessment Management System for Interactive Learning. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 284–289.

[20] H. Kubrak, M. Ehlers, K. Piecha, T. Walcher, G. Braun, and P. Prade. 2023. EXaHM–Application Oriented, Digital EXamination System at Hochschule München. *Electronic Journal of e-Learning* 21, 4 (2023), 328–334.

[21] Matti Kuikka, Markus Kitola, and Mikko-Jussi Laakso. 2014. Challenges when introducing electronic exam. *Research in Learning Technology* 22, 0 (2014).

[22] Oka Kurniawan, Norman Tiong Seng Lee, and Christopher M. Poskitt. 2020. Securing Bring-Your-Own-Device (BYOD) Programming Exams. In *Proceedings of the 51st Technical Symposium on Computer Science Education (SIGCSE)*. ACM, 880–886.

[23] Bastian Küppers. 2021. *Development of a framework for e-assessment on students' devices: on the challenges of applying BYOD in digital examinations.* Ph. D. Dissertation. RWTH Aachen.

[24] B. Küppers and U. Schroeder. 2016. Bring Your Own Device For E-Assessment - A Review. In *8th International Conference on Education and New Learning Technologies*. IATED.

[25] Bastian Küppers and Ulrik Schroeder. 2020. Teacher's Perspective on e-Assessment: A Case Study from Germany. In *Proceedings of the 12th International Conference on Computer Supported Education - Volume 1: CSEDU*. SciTePress.

[26] Matthias Linhuber, Jan Philip Bernius, and Stephan Krusche. 2024. Constructive Alignment in Modern Computing Education: An Open-Source Computer-Based Examination System. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research (Koli Calling '23)*. ACM.

[27] Gosia Migut, Dennis Koelma, Cees G. M. Snoek, and Natasa Brouwer. 2018. Cheat Me Not: Automated Proctoring of Digital Exams on Bring-Your-Own-Device. In *Proceedings of the 23rd Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE)*. ACM, 388.

[28] Andrea Nardi and Maria Ranieri. 2019. Comparing paper-based and electronic multiple-choice examinations with personal devices: Impact on students' performance, self-efficacy and satisfaction. *British Journal of Educational Technology* 50, 3 (2019), 1495–1506.

[29] Philip M. Newton and Keioni Essex. 2023. How Common is Cheating in Online Exams and did it Increase During the COVID-19 Pandemic? A Systematic Review. *Journal of Academic Ethics* (2023).

[30] Chris Piech and Chris Gregg. 2018. BlueBook: A Computerized Replacement for Paper Tests in Computer Science. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. ACM, 562–567.

[31] Teemu Rajala, Erkki Kaila, Rolf Lindén, Einari Kurvinen, Erno Lokkila, Mikko-Jussi Laakso, and Tapio Salakoski. 2016. Automatically Assessed Electronic Exams in Programming Courses. In *Proceedings of the Australasian Computer Science Week Multiconference (ACSW)*. ACM.

[32] Per Runeson and Martin Höst. 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering* 14, 2 (2009), 131–164.

[33] Guttorm Sindre and Aparna Chirumamilla. 2015. E-exams versus paper exams: A comparative analysis of cheating-related security threats and countermeasures.

[34] Ben Stephenson. 2018. An experience using on-computer programming questions during exams. In *Proceedings of the 23rd Western Canadian Conference on Computing Education*. 1–6.

[35] Craig Zilles, Matthew West, David Mussulman, and Carleen Sacris. 2018. Student and Instructor Experiences with a Computer-Based Testing Facility. In *10th International Conference on Education and New Learning Technologies*. 4441–4450.

[36] Małgorzata Ćwil. 2019. Teacher's Attitudes towards Electronic Examination—a Qualitative Perspective. *International Journal of Learning and Teaching* 5, 1 (2019), 77–82.

[5] https://osf.io/83zpf/?view_only=4a7f0be384414d14be72fa5998908f59