Asian Journal of Distance Education

The Examination of Online and Paper-Pencil Test Scores of Students Engaged in Online Learning

Necati Taşkın, Kerem Erzurumlu

Abstract: In this study, online test scores and paper-pencil test scores of students studying through online learning were examined. Causal-comparative research was used to determine the distribution of students' test scores and to examine the relationship between them. The participants of the research are freshman students studying in 12 faculties and 8 colleges of a state university in Türkiye. The distribution of students' test scores is depicted by means, standard deviation, percentage, and graphs. The correlation coefficient was examined to find and interpret the amount of relationship between the test scores of the students. According to the findings, it was seen that the online test scores of the students were higher than the paper-pencil test scores. At the same time, it was observed that the passing of the course rates in online test exams was higher than in the paper-pencil test exams. It was observed that the relationship between the paper-pencil test scores of the students and the online test scores was lower than the relationship between the paper-pencil test scores and the paper-pencil test scores. There is an inconsistency between students' paper-pencil test scores and online test scores. The rise in students' online test scores to un-proctored online exams as the reason for the inconsistency. Moving online exams to proctored exam environments, using computerized adaptive testing, or including online activities in the assessment may reduce this inconsistency.

Keywords: online learning, online assessment, un-proctored exam, online test, paper-pencil test

Highlights

What is already known about this topic:

- The demand for online assessment methods is steadily rising in education.
- Online exams are susceptible to cheating.
- Concerns are growing about the accuracy of the online assessment.

What this paper contributes:

- It will show the relationship between the students' online test scores and the paper-pencil test scores.
- It will provide valuable information about the functionality of un-proctored online testing.

Implications for theory, practice and/or policy:

- It will help universities and educators develop innovative solutions for online exams.
- It will raise awareness among educators about creating a safe online exam environment.

Published by Asian Society for Open and Distance Education (ASODE), Japan
ISSN 1347-9008 http://www.asianjde.com/

Introduction

With the increasing interest in online learning, online exams have started to be used more as an assessment method. The Covid-19 pandemic has further increased the importance of online assessment (Daniel, 2020; Rahim, 2020), showing that online exams will be an important part of education in the future (Langenfeld, 2020). Online exams reduce the workload of instructors, simplify the process, reduce costs, and provide flexibility in the context of time and space to learners (Butler-Henderson & Crawford, 2020). Although online exams offer more opportunities than paper-pencil exams (Born et al., 2019; Zhang et al., 2019), they may cause reliability issues (Palloff & Pratt, 2009) and cheating (Hylton et al., 2016).

Un-proctored tests are often the first choice of instructors for online assessment (Clark et al., 2020). In these tests, students are not supervised by a live supervisor or technologies such as artificial intelligence during the exam. So, some security measures are taken, such as time limitation, random ordering of questions and options, creating a question pool, limiting the test time, and delaying feedback (Şenel & Şenel, 2021). Although these measures reduce the possibility of cheating, they do not guarantee its security. The biggest danger that threatens security is that students receive help from anyone other than themselves (Kılınç et al., 2021). This raises concerns about the accuracy and fairness of online assessment results.

Measurement and evaluation are some of the biggest problems experienced in the online learning (Bozkurt, 2020). There are many issues that need to be investigated regarding the consistency, validity, reliability, and whether they allow for better measurement or higher student achievement of online exams (Butler-Henderson & Crawford, 2020). Examinations provide statistical data on the evaluation process as well as deciding on student proficiency. These data provide many opportunities such as improving the quality of exams, evaluating the process and improving measurement accuracy (Goldhammer et al., 2020). In this context, the usability of online exams is an important point that should be emphasized (Solak et al., 2020). The contamination of evaluation by undesirable behaviors such as cheating raises concerns that online exams do not show real student success (Ural & Takaoğlu, 2023). Therefore, students' achievements should be evaluated by comparing online and paper-pencil tests (Solak et al., 2020). The statistical analysis of the data obtained from the students will contribute to the development of online assessment. It will also guide future studies on the functionality and usability of un-proctored online tests.

Literature

Theoretical background on the research problem is given under the sub-headings of assessment, online assessment, and un-proctored online exams.

Assessment

Assessment is a basic component of the curriculum (Ilgaz & Afacan Adanır, 2020). Assessment is defined as the process of collecting data about students' performance, analyzing them, and producing results (Gülbahar, 2016). It is used to decide the proficiency of students at the end of a teaching-learning process (Zacharis, 2010). Students take many exams throughout the year to be successful in a course, pass the class and get a degree. Multiple-choice test exams are the widely used method to determine students' proficiency (Dennick et al., 2009). Multiple-choice tests are objective results, provide the opportunity to evaluate learning outcomes at various levels, and are easy to create and grade (Nguyen et al., 2020). Since each item has a fixed score in these tests, students may predict their scores and their learning situations are evaluated objectively (Brown, 2019). So, multiple-choice tests are generally the first choice of instructors and universities in the evaluation of students' proficiency (Clark et al., 2020).

Online Assessment

The online assessment offers significant advantages to both students and instructors such as instant feedback, reusability, ease of storage, statistical data, rich evaluation tools, personalized sessions, ease of evaluation, and resource savings (Şenel & Şenel, 2021). Online exams are more economical than paper-pencil exams (Attia, 2014). It allows instructors to spend less time grading students' assignments. (Dennick et al., 2009). In addition, statistical analysis is easily accessible (Clark et al., 2020). Most importantly, it offers students flexibility in the context of time and space (Zhang et al., 2019). In the literature, it is seen that students have a positive attitude toward online exams (Ilgaz & Afacan Adanır, 2020). Students prefer online exams instead of paper-pencil exams because they add value to their learning, provide instant feedback, offer self-assessment opportunities (Cabı, 2016; Sorensen, 2013), are quick-announced of students' exam scores, and provide ease of rearranging answers (Jawaid et al., 2014; Pagram et al., 2018). Although online exams offer significant advantages, their biggest disadvantage is that they are more vulnerable to academic dishonesty (Clark et al., 2020; Fask et al., 2014). The flexibility of online exams makes it easy for students to cheat on these exams (Attia, 2014; Fask et al., 2014; Pagram et al., 2018). For this reason, some measures should be taken in online exams to minimize the risks.

Live and artificial intelligence-controlled mechanisms are used to reduce the undesirable behavior of students, but the use of these high-tech structures is not common (Hylton et al., 2016; Karim et al., 2014). So, un-proctored online exams are generally preferred as it is an easy to implement and a cost-effective method (Arthur et al., 2010; Gibby, 2009). Un-proctored online exams are techniques in which students are not supervised during the exam by a live supervisor or technology-based approaches such as deep learning, artificial intelligence, and machine learning (Milone et al., 2017). Students state that un-proctored online exams make cheating easier, and they are concerned about this issue (Cabı, 2016; Pagram et al., 2018). Situations such as joining another person in the exam, using an instant messaging program, making a phone connection, and making a remote desktop connection are some of the cheating threats in online exams (Ullah et al., 2016). Bloemers et al. (2016) stated that students violated the exam rules by getting help from others, searching the internet, and sharing test questions. Srikanth and Asmatulu (2014) reported in their study that 70% of students admitted to cheating in online exams. On the other hand, Rios and Liu (2017) state that un-proctored online exams can be used to draw inferences about students' learning outcomes. Some studies in the literature seem to recommend un-proctored as an alternative to the paper-pencil test (Prisacari, 2017; Ural & Takaoğlu, 2023). It is unclear in the literature regarding the usability and validity of un-proctored tests. However, instructors have the responsibility of creating a secure exam environment and administering a valid exam. Ensuring valid and reliable assessment is an important issue for researchers and educators (Adzima, 2020). Therefore, the discussions about un-proctored online exams should be resolved as soon as possible.

Un-proctored Online Exam

Some measures are taken to prevent cheating in un-proctored online exams, such as time limit, random question selection, random ordering of questions and/or options, creating a question pool, turning on the webcam, and blocking the web tabs (Kılınç et. al., 2021; Şenel & Şenel, 2021). In addition, measures such as limiting the number of access, limiting the test time, and delaying feedback are used to prevent cheating (Rahim, 2020). Langenfeld (2020) recommends publishing the test to all students at the same time. It also suggests a short test response time, one-time access to the test, and no return to the answered question. The use of a question pool significantly reduces the possibility of cheating (Clark et al., 2020). It is tried to prevent resource research by limiting the response time of the test. Given enough time and opportunity, students will research answers from different sources and share answer among them. Therefore, measures such as limiting test time and randomly selecting questions are necessary. The most commonly used method in online exams to prevent cheating is to randomly distribute questions from the question pool. The width of the question

pool increases the reliability of the exam (Kırmacı & Kılıç Çakmak, 2020). Kılınç et al. (2021) suggest increasing the number of questions in the question pool for the usability of test exams. By establishing an effective test design mechanism through multiple question banks, the risks of cheating can be reduced (Sullivan, 2016). However, even all these measures will not guarantee online test security.

The Importance of Study

Assessment is a fundamental element of online learning, as it is in face-to-face education. Accurate measurement results are needed to assess the quality of learning and the effectiveness of the training program. Online assessment can lead to an unfair distribution of test scores among students in the assessment of learning outcomes. The expansion of online assessment has raised important questions about their ability to accurately reflect true student success (Garg & Goel, 2022). While online assessments offer certain advantages, such as flexibility and rapid feedback, they also present challenges related to cheating, assessment format, and more. The effectiveness of online assessments depends on their design, implementation, and careful evaluation of their results to provide a fair and accurate measure of student achievement. Butler-Henderson and Crawford (2020) state that universities have been slow to develop innovative solutions, although they have started to use the opportunities provided by online exams. This study will contribute to educators, administrators, and policymakers by offering important insights into the effective development and use of online exams. This can support the creation of more reliable and valid assessment tools and improve our understanding of student success.

Purpose of the Research

This study aims to examine the paper-pencil test scores and un-proctored online test scores of students studying through online learning. The research questions to be answered under this main objective are as follows.

(Q1) What is the distribution of students' paper-pencil and online exam scores?

(Q2) What is the course passing rate of the students according to the exam type?

(Q3) Is there a significant relationship between the students' fall and spring term mean scores?

a) Is there a significant relationship between the mean scores of the students in the same exam type?

b) Is there a significant relationship between the mean scores of students from different types of exams?

Methodology

In this study, casual-comparative research was conducted. It is a type of research aimed at determining the causes of an existing situation, the variables affecting this cause, or the consequences of this effect (Fraenkel et al., 2018). There is no intervention for students. The researcher has no influence on the formation of student groups.

Participant

The participants of the research consist of students studying at 12 faculties and 8 colleges of a state university and they are learning History and Foreign Language courses through online learning. Only freshman students who joined all the midterm exam and final exams in both the spring and fall semesters were included in the study. In this context, the research was carried out on 12489 students who took History course and 11917 students who took Foreign Language courses between 2017 and 2021. The distribution of participant numbers by years and courses is shown in Table 1.

Table 1. Distribution of participants

Course / Year	History	Foreign Language
	N	N
2017-2018	3727	4247
2018-2019	3004	3538
2019-2020	2833	1299
2020-2021	2925	2833
Total	12489	11917

N: Number of students

Research Procedures

Students received their education through online learning. The weekly course document prepared in "*pdf*" format was shared on the MOODLE learning management system. In addition, live courses (synchronous) were held on BigBlueButton every week. Live courses were recorded, and students were able to access live course recordings whenever and wherever they wanted. Students took the midterm exam in the 8th week and the final exam in the 16th week of the 16-week education period. Students were evaluated with exams in different types in different periods (Table 2).

Table 2. Types of exams

Course / Year	History		Foreign Language	
	Fall	Spring	Fall	Spring
2017-2018	Pape-Pencil Test	Pape-Pencil Test	Pape-Pencil Test	Pape-Pencil Test
2018-2019	Pape-Pencil Test	Pape-Pencil Test	Pape-Pencil Test	Pape-Pencil Test
2019-2020	Pape-Pencil Test	Oline Test	Pape-Pencil Test	Oline Test
2020-2021	Oline Test	Oline Test	Oline Test	Oline Test

Paper-pencil tests were held at the same time across the entire campus. In each course, students were given question booklets containing multiple-choice tests with 20 questions. These tests were administered under the supervision of an instructor. Students marked their answers on optical forms provided for the exams. Students' answers were evaluated through optical readers.

Online exams were conducted un-proctored. Students joined the exams with their own usernames and passwords through the learning management system. The online test consists of 20 multiple-choice test questions for each course. Students were granted access to the exams for duration of 24 hours. Upon joining the session, students were allotted 20 minutes to respond to the questions. Each student had a single opportunity to enter the exam. The questions were presented to the students sequentially, with no option to revisit answered questions. Question choices and questions were shuffled for each student. The browser tab was restricted during the session.

Data Collection Tools

To determine the academic achievement of the students, 20-question multiple-choice tests were used. Each question has four options and one correct answer. A question answered incorrectly does not affect the number of correct answers. Each correctly answered question has a value of 5 points and a maximum of 100 points is taken from the exam. 40% of midterm exam scores and 60% of final exam scores constitute the semester means. Students who score 60 or more in a course are considered successful and pass the course.

The same question pool was used for both paper-pencil and online tests. The question pool has been created by the course instructors through many years of work. Each question in the question pool has statistical data based on previous exam practices. Tests were created automatically, taking into account the subject scope and difficulty level of the questions. To ensure the content validity of the tests, categories were created according to topics. For example, the question distribution for the fall semester exams of the history course is given in Table 3.

Table 3. History course fall semester exam question distribution

Topics	Number of Questions in the Pool	Midterm Exam Test	Final Exam Test
Basic concepts like Revolution	36	3	1
The structure of the Ottoman Empire	40	3	1
The Ottoman geopolitics and foreign policy	41	3	1
World War I	45	3	1
Mondros Armistice	39	2	1
Mustafa Kemal's opinions and his pass to Anatolia	40	2	1
Occupation of Istanbul	34	2	1
First activities of the Grand National Assembly of Turkey	36	2	1
National Struggle Fronts (South and Southeast)	14	-	2
National Struggle Fronts (West)	39	-	2
Treaty of Sevres	34	-	2
Mudanya Armistice	47	-	2
Lozan Peace Conference	45	-	2
Foreign policy in the Ataturk period	26	-	2

Effective test design is achieved through the question pool (Sullivan, 2016) and injustices among students can be eliminated (Sorensen, 2013). According to Kırmacı & Kılıç Çakmak, (2020) use of the question pool increases the reliability of the test exam. Kılınç et al. (2021) state that equivalent tests can be prepared using the question pool. In this context, equivalent tests were created in terms of the validity and reliability of the tests by using the question pool. For example, the statistical data of the history course midterm exam are shown in Table 4.

Table 4. Statistical information of the history test

Question	pj	rjx	sj	rj
1	0.64	0.45	0.47	0.30
2	0.51	0.40	0.49	0.25
3	0.76	0.36	0.42	0.32
4	0.60	0.53	0.48	0.29
5	0.59	0.55	0.49	0.29
6	0.70	0.42	0.45	0.32
7	0.71	0.41	0.45	0.32
8	0.62	0.41	0.48	0.30
9	0.38	0.51	0.48	0.18
10	0.51	0.48	0.49	0.25
11	0.64	0.47	0.47	0.30
12	0.41	0.52	0.49	0.20
13	0.72	0.31	0.44	0.32
14	0.66	0.48	0.47	0.31
15	0.41	0.42	0.49	0.20
16	0.48	0.30	0.49	0.24

pj: Item difficulty index; rjx Discrimination index; sj: Standard deviation; rj: Item reliability coefficient

The mean of this test given as an example is 11.62, the standard deviation is 3.74, the variance is 13.95, the average difficulty is 0.56 and the Kr-20 reliability coefficient is 0.71. The average difficulty of the test is between 0.40-0.60, which is considered ideal (Başol, 2018). The reliability coefficient of the test is above the acceptable value (Kr-20>70). The discrimination power of the test items is above 0.30, which is considered good (Başol, 2018). All tests are medium-difficulty, acceptable reliability coefficient, and similar statistical values.

Data Analysis

The mean and standard deviation values were used to describe the test scores of the students. Bar graphs are used to make the distributions more understandable. The correlation coefficient was used to find and interpret the amount of relationship between the test scores of the students. Since the mean scores of the students showed a continuous and normal distribution (Table 5), the Pearson correlation coefficient was used to explain the relationship between the variables (Büyüköztürk, 2022).

Table 5. Skewness and kurtosis values regarding mean scores

Course	2017-2018				2018-2019				2019-2020				2020-2021			
	Fall		Spring		Fall		Spring		Fall		Spring		Fall		Spring	
	S	K	S	K	S	K	S	K	S	K	S	K	S	K	S	K
History	0.15	0.52	0.13	0.22	0.12	0.64	0.12	0.24	0.11	0.53	0.58	0.32	0.55	0.11	0.73	0.55
Foreign Language	0.47	0.18	0.52	0.53	0.23	0.50	0.48	0.33	0.30	0.56	0.71	0.15	1.01	0.90	0.03	0.01

S: Skewness, K: Kurtosis

As can be seen in Table 5, since the skewness and kurtosis values of the students' average scores in the fall and spring semesters are in the range of ±1.5, the assumption of normality is met (Tabachnick & Fidell, 2019).

Limitations

The groups compared in this study consist of the same students. At the same time, groups trained with the same teaching method. However, comparing students' midterm and final exam scores is a limitation of this study. The participants of this study were first-year students of a public university in Türkiye. Results may vary for students from different cultures and educational levels. In addition, the online education provided to students is limited to the capabilities of the MOODLE learning management system and the BigBlueButton synchronous course module.

Findings

Exam Score Distribution of Students

The midterm, final, and semester mean scores of the students for the History course are shown in Table 6.

Table 6. History course midterm, final, and semester scores

Course	Fall						Spring					
	Midterm		Final		Semester		Midterm		Final		Semester	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
2017-2018	58.4	18.41	54.56	15.94	56.1	15.21	60,82	14.33	61	16.32	60,93	13.69
2018-2019	59.61	17.37	57.65	17.88	58.44	15.96	59,1	15.11	61.45	16.04	60.51	13.95
2019-2020	63.31	17.86	64.01	16.70	63.73	15.36	73.78*	15.12*	77,04*	14.33*	75.74*	12.17*
2020-2021	72.65*	16.28*	74.80*	15.47*	73.94*	13.22*	78.81*	13.58*	79.05*	13.46*	78.95*	11.17*

* Online exam

When Table 6 is examined, it is seen that the means scores of the students in the online tests are higher than in the paper-pencil tests. For example, while the student's midterm exam mean score in the fall semester of the 2019-2020 academic year is 63.31 and the final exam mean score is 64.01, it is seen that the online midterm test mean score in the spring semester is 73.78 and the final test mean score is 77.04. Likewise, in the 2017-2018 and 2018-2019 academic years, the midterm and final test mean scores made with paper and pencil ranged between 54.56 and 60.82, while the online midterm and final test mean scores in the 2020-2021 academic year varied between 72.65 and 79.05. The students' mean score in the fall and spring semesters of the History course is higher in the online tests. The midterm, final, and semester mean scores of the students for the Foreign Language course are shown in Table 7.

Table 7. Foreign language course midterm, final, and semester scores

Year	Fall						Spring					
	Midterm		Final		Semester		Midterm		Final		Semester	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
2017-2018	53.88	18.31	47.33	18.38	49.95	16.27	49.82	19.16	48.65	23.23	49.12	19.16
2018-2019	54.82	18.54	54.01	18.38	54.34	16.48	51.16	20.33	48.50	18.00	49.56	17.03
2019-2020	51.83	17.59	58.48	17.95	55.82	17.08	73.91*	21.98*	76.55*	19.98*	75.49*	17.70*
2020-2021	83.63*	14.23	80.73*	15.51*	81.89*	12.70*	80.22*	17.39*	78.71*	18.66*	79.31*	14.29*

* Online exam

It is seen that students' mean score of the online Foreign Language test is higher than the mean score of the paper-pencil test. While the means of the paper-pencil test scores of the student varies between 47.33 and 54.82, it is seen that the mean of the online test scores varies between 73.91 and 83.63.

Students' Grade Pass Rates

The rates of students passing the History and Foreign language courses between the academic years of 2017-2018 and 2020-2021 are given in Table 8.

Table 8. Course pass rates

Year	History				Foreign Language			
	Fall		Spring		Fall		Spring	
	n	%	n	%	n	%	n	%
2017-2018	1492	40.03	2045	54.87	1103	45.08	1229	50.22
2018-2019	1387	46.17	1585	52.76	1279	36.16	943	26.66
2019-2020	1663	58.70	2554*	90.15*	509	39.18	1050*	80.83*
2020-2021	2491	85.16*	2766*	94.56*	2441*	86.16*	2426*	85.63*

* Online exam, n: Number of students passing the course

When Table 8 is examined, the passing rate of course in online test is higher than in paper-pencil tests. While the pass rate of the students in the Foreign Language course paper-pencil exams varies between 26.6% and 50.22%, it is seen that this rate varies between 80.83% and 86.16% in the online tests.

The Relationship between Same Exam Types

Table 9. The relationship between students' paper-pencil exam scores

Course	Semester	2017-2018			2018-2019		
		n	r	p	n	r	p
History	Fall Spring	3726	0.669	0.000	3004	0.700	0.000
Foreign Language	Fall Spring	4247	0.643	0.000	3537	0.722	0.000

It is seen that there is a moderate, positive, and significant correlation (Büyüköztürk, 2022) between the students' mean scores in the paper-pencil test in the fall and spring semesters of the 2017-2018 academic year ($0.3<r<0.7$, $p<0.01$). It is seen that there is a high level of correlation (Büyüköztürk, 2022) between the fall and spring semester mean scores of the 2018-2019 academic year ($r>0.7$, $p<0.01$).

Table 10. The relationship between students' online exam scores

Course	Semester	2020-2021		
		n	r	p
History	Fall Spring	2925	0.501	0.000
Foreign Language	Fall Spring	1689	0.405	0.000

It is seen that there is a moderate, positive and significant correlation between the average scores of the students in the online exams in the fall and spring semesters of the 2020-2021 academic year ($0.3<r<0.7$, $p<0.01$). It is seen that the level of correlation between the students' online exam mean scores (Table 10) is lower than the correlation level between paper-pencil exam mean scores (Table 9).

The Relationship between in Different Exam Types

Table 11. The relationship between students' means scores in different exam types

Course	Semester	2019-2020		
		n	r	p
History	Fall (Paper-pencil) Spring (Online)	2832	0.338	0.000
Foreign Language	Fall (Paper-pencil) Spring (Online)	1299	0.159	0.000

It is observed that there is a moderate, positive and significant correlation between the students' History course paper-pencil test score means in the fall semester and the spring semester online test score means ($0.3<r<0.7$, $p<0.01$). On the other hand, it is seen that there is a low-level correlation between the Foreign language course test score means ($r<0.3$, $p<0.01$). It is seen that the correlation between the mean scores of the students in different exam types (Table 11) is lower than the level of same exam type (Table 10 and Table 9). In fact, it is seen that the correlation between the mean scores in different types of exams is lower than the correlation between exams with paper-pencil.

Discussion

This study aimed to examine the paper-pencil and online exam scores of students studying through online learning. While the mean scores of the students for the same exam type were close to each other, it was seen that the mean score of the online exam was higher than the mean of the paper-pencil exam (1). Another finding is that the course passing rate of students in online exams is higher than in paper-pencil exams (2). The most important finding is that the correlation level between the students' online exam mean scores and the paper-pencil exam mean score is lower than the mean score in the same exam type (3). The findings of this study are discussed within the framework of other studies in the literature.

The first finding of this study is that students' online exam mean scores are higher than their paper-pencil exam mean scores. Similarly, the course passing rates of students in online exams have increased compared to paper-pencil exams. Bloemers et al. (2016) and Steger et al. (2020) found in their study that the distribution of scores in un-proctored online exams is high. Although Orr and Foster (2013) stated that students were more successful in online exams than in pencil-paper exams, Alessio et al. (2018) stated that this increase in scores was due to the fact that students turned to different ways to find answers to the questions in the exams. Similarly, Ilgaz and Afacan Adanır (2020) state that students' performance differs according to the format of the exam (paper-pencil or online), and they state that this may be due to the fact that students apply to different sources in online exams. Bloemers et al. (2016) state that students violate the rules by getting help from their friends, searching the internet, and sharing questions in online exams. In online exams, students are more likely to cheat because they can access different resources more easily (Ravasco, 2012). The study of Garg and Goel (2022) also reveals that academic dishonesty is a common problem in online education and leads to higher test scores of students.

The vulnerability of un-proctored online exams to academic misconduct may be the main reason for high student exam scores and increased course pass rates. Because, while it is seen in the literature that students' proctored online exam scores are generally at the same level as pencil-paper exam scores (Ardid et al. 2015; Berkey & Halfond, 2015; Boevé et al., 2015; Karay et al., 2015; Karim et al.,

2014; Kemp & Grieve 2014; Lilley et al., 2016; Weiner & Hurtz, 2017), it is that the scores in the un-proctored online exam increase (Alessio et al., 2018; Bloemers et al., 2016; Ilgaz & Afacan Adanır, 2020; Steger et al., 2020). The finding in this study that the level of correlation between the online and paper-and-pencil exam mean scores is low also supports this situation. In the literature, it is seen that proctored online exams are more successful in minimizing academic misconduct, creating a safer exam environment, and achieving a balanced score distribution (Hylton et al., 2016; Karim et al., 2014; Weiner & Hurtz, 2017). We are seeing increasing use of artificial intelligence and machine learning techniques to automate the detection of dishonest behavior (Garg & Goel, 2022). Considering that online learning and online assessment will be an important part of education in the future, it can be said that these techniques will play an important role in preventing undesirable behaviors.

The most striking finding reached in this study is that while there is a high correlation between the students' paper-pencil exam scores, there is a lower correlation between different exam types' mean scores. This finding shows that there is an inconsistency between the students' un-proctored online exam scores and their paper-pencil exam scores. Although studies in the literature point to un-proctored online exams as the main reason for this inconsistency (Garg & Goel, 2022), it will continue to be used as it is an easy and cost-effective method (Fask et al., 2014; Hylton et al., 2016). So, there is a need to develop new solutions to prevent academic misconduct in un-proctored online exams. Perhaps online assessment should not be limited to exams that only measure students' knowledge. Goldhammer et al. (2020) state that other online activity data such as content download, length of stay in the system, and number of posts on the discussion forum will also be used to determine academic achievement in the future. In some studies, it is seen that online learning experiences are used to predict the academic achievement of students (Goh et al., 2017; O'Shea et al., 2015). Clark et al. (2020) state that it is wrong to use the same question type used in face-to-face exams in online exams. Instead of standardized tests consisting of fixed-score questions (Brown, 2019), adapted tests in which questions are arranged according to student's performance can be used for online exams. Adapted tests reveal more reliable and valid results than standard tests (Liu et al., 2019). In summary, using adapted test questions and/or incorporating online activities into assessment can help achieve more consistent test scores.

Conclusion and Suggestions

In this study, it was seen that the scores of the students in the online exams were higher than the paper-pencil exams. At the same time, passing rates in online exams are quite high compared to paper-pencil exams. There is an inconsistency between students' paper-pencil test scores and online test scores. These results indicate that un-proctored online exams do not reflect real student achievement. Although it is stated in the literature that this situation is due to the vulnerability of un-proctored online exams to cheating, the findings obtained within the scope of this study do not clearly reveal this conclusion. However, it can be said that students' paper-and-pencil exam results are not equivalent to un-proctored online exam results.

Incorrect evaluations can lead to unfair outcomes. Students who deserve higher grades may be overlooked, while those who receive undeserved grades may enjoy unwarranted benefits. If students are not evaluated accurately, they may focus on memorization or shortcuts to achieve good grades rather than genuine learning. When diplomas and certificates do not accurately represent a student's abilities, it undermines the credibility of these credentials. Online education has become increasingly popular, but employers may question the competence of online graduates. Incorrect evaluations can also affect research and scientific studies. To combat these issues, it is imperative that comprehensive solutions are developed promptly.

Although the literature points to live or artificial intelligence-monitored mechanisms to obtain more balanced exam score distributions, the cost of these mechanisms limits their use. The fact that the

proctored online exam mechanisms are not yet at the desired level shows that the use of un-proctored online exams will continue for a while. Educators should find solutions to this problem experienced in online assessment as soon as possible. A safe exam environment should be created to maintain the balance between students' exam scores.

- Students' final achievements at the end of online learning can be obtained through multiple assessments (presentations, e-portfolios, group projects, etc.). Including un-proctored online exams with a certain weight in the evaluation can minimize their misleading effects.
- Including students' online activities such as synchronous course attendance and content review rate in the evaluation may provide a more balanced scoring.
- Using adapted test questions instead of standard tests in online exams may produce more reliable and balanced results.

References

Adzima, K. (2020). Examining online cheating in higher education using traditional classroom cheating as a guide. *Electronic Journal of E-Learning, 18*(6), 476-493. https://doi.org/10.34190/JEL.18.6.002

Alessio, H. M., Malay, N., Maurer, K., Bailer, A. J., & Rubin, B. (2018). Interaction of proctoring and student major on online test performance. *The International Review of Research in Open and Distributed Learning, 19*(5). https://doi.org/10.19173/irrodl.v19i5.3698

Ardid, M., Gómez-Tejedor, J. A., Meseguer-Dueñas, J. M., Riera, J., & Vidaurre, A. (2015). Online exams for blended assessment. Study of different application methodologies. *Computers & Education, 81*, 296–303. https://doi.org/10.1016/j.compedu.2014.10.010

Arthur, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment, 18*(1), 1–16. https://doi.org/10.1111/j.1468-2389.2010.00476.x

Attia, M. (2014). Postgraduate students' perceptions toward online assessment: The case of the faculty of education, Umm Al-Qura University. In A. Wiseman, N. Alromi, & S. Alshumrani (Eds.), *Education for a Knowledge Society in Arabian Gulf Countries* (pp. 151–173). Emerald Group Publishing Limited. https://doi.org/10.1108/s1479-3679_2014_0000024015

Başol, G. (2018). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi.

Berkey, D., & Halfond, J. (2015). *Cheating, student authentication and proctoring in online programs*. New England Journal of Higher Education.

Bloemers, W., Oud, A., & van Dam, K. (2016). Cheating on unproctored internet intelligence tests: Strategies and effects. *Personal Assessment and Decisions, 2*(1), 3, 21–29. https://doi.org/10.25035/pad.2016.003

Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). Introducing computer-based testing in high-stakes exams in Higher Education: Results of a field experiment. *PLoS One, 10*(12), e0143616. https://doi.org/10.1371/journal.pone.0143616

Born, S., Fink, A., Spoden, C., & Frey, A. (2019). Evaluating different equating setups in the continuous item pool calibration for computerized adaptive testing. *Frontiers in Psychology, 10*, 1277. https://doi.org/10.3389/fpsyg.2019.01277

Bozkurt, A. (2020). Koronavirüs (Covid-19) pandemi süreci ve pandemi sonrası dünyada eğitime yönelik değerlendirmeler: Yeni normal ve yeni eğitim paradigması. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi, 6*(3), 112-142. https://dergipark.org.tr/en/pub/auad/issue/56247/773769

Brown, B. (2019). Negative effects of standardized testing [Capstone Projects and Master's Theses], 449. California State University. https://digitalcommons.csumb.edu/caps_thes_all/

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education, 159*, 104024. https://doi.org/10.1016/j.compedu.2020.104024

Büyüköztürk, Ş. (2022). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi Yayıncılık.

Cabı, E. (2016). The perception of students on e-assessment in distance education. *Journal of Higher Education & Science, 6*(1), 94–101. https://dergipark.org.tr/en/pub/higheredusci/issue/61489/918150

Clark, T. M., Callam, C. S., Paul, N. M., Stoltzfus, M. W., & Turner, D. (2020). Testing in the time of COVID-19: A sudden transition to unproctored online exams. *Journal of Chemical Education, 97*(9), 3413-3417. https://doi.org/10.1021/acs.jchemed.0c00546

Daniel, S. J. (2020). Education and the COVID-19 pandemic. *PROSPECTS, 49*(1), 91-96. https://doi.org/10.1007/s11125-020-09464-3

Dennick, R., Wilkinson S., & Purcell N. (2009) Online eAssessment: AMEE Guide No. 39, *Medical Teacher, 31*(3), 192-206. https://doi.org/10.1080/01421590902792406

Fask, A., Englander, F., & Wang, Z. (2014). Do online eams facilitate cheating? An experiment designed to separate possible cheating from the effect of the online test taking environment. *Journal of Academic Ethics, 12*, 101–112. https://doi.org/10.1007/s10805-014-9207-1

Fraenkel, J., Wallen, N., & Hyun, H. (2018). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.

Garg, M., & Goel, A. (2022). A systematic literature review on online assessment security: Current challenges and integrity strategies. *Computers & Security, 113*, 102544. https://doi.org/10.1016/j.cose.2021.102544

Gibby, R. E., Ispas, D., McCloy, R. A., & Biga, A. (2009). Moving beyond the challenge to make unproctored internet testing a reality. *Industrial and Organizational Psychology, 2*(1), 64–68. https://doi.org/10.1111/j.1754-9434.2008.01110.x

Goh, C., Leong, C., Kasmin, K., Hii, P., & Tan, O. (2017). Students' experiences, learning outcomes and satisfaction in e-Learning. *Journal of e-Learning and Knowledge Society, 13*(2). https://www.learntechlib.org/p/188116/

Goldhammer, F., Scherer, R., & Greiff, S. (2020). Advancements in technology-based assessment: Emerging item formats, test designs, and data sources. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.03047

Gülbahar, Y. (2016). E-Değerlendirme. K. Çağıltay & Y. Göktaş (Eds.), Öğretim Teknolojilerinin Temelleri: Teorileri, Araştırmalar, Eğilimler içinde (s. 655-666). Pegem Akademi.

Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education, 92*, 53-63. https://doi.org/10.1016/j.compedu.2015.10.002

Ilgaz, H., & Afacan Adanır, G. (2020). Providing online exams for online learners: Does it really matter for them? *Education and Information Technologies, 25*(2), 1255-1269. https://doi.org/10.1007/s10639-019-10020-6

Jawaid, M., Moosa, F. A., Jaleel, F., & Ashraf, J. (2014). Computer based assessment (CBA): Perception of residents at Dow University of Health Sciences. *Pakistan Journal of Medical Sciences, 30*(4), 688–691. https://doi.org/10.12669/pjms.304.5444

Karay, Y., Schauber, S. K., Stosch, C., & Schüttpelz-Brauns, K. (2015). Computer versus paper—Does it make any difference in test performance? *Teaching and Learning in Medicine, 27*(1), 57–62. https://doi.org/10.1080/10401334.2014.979175

Karim, M. N., Kaminsky, S. E., & Behrend, T. S. (2014). Cheating, reactions, and performance in remotely proctored testing: An exploratory experimental study. *Journal of Business and Psychology, 29*(4), 555-572. https://doi.org/10.1007/s10869-014-9343-z

Kemp, N., & Grieve, R. (2014). Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. *Frontiers in Psychology, 5*, 1278. https://doi.org/10.3389/fpsyg.2014.01278

Kılınç, H., Okur, M. R., & İlker, U. S. T. A. (2021). The opinions of field experts on online test applications and test security during the covid-19 pandemic. *International Journal of Assessment Tools in Education, 8*(4), 975-990. https://doi.org/10.21449/ijate.875293

Kırmacı, Ö., & Kılıç Çakmak, E. (2020). Çevrimiçi Öğrenme Ortamlarında Değerlendirme. In E. Kılıç Çakmak & S. Karataş (Eds.), *Çevrimiçi Öğrenme: Farklı Bakış Açıları* (s. 387-419). Pegem Akademi.

Langenfeld, T. (2020). Internet-based proctored assessment: Security and fairness issues. *Educational Measurement: Issues and Practice, 39*(3), 24-27. https://doi.org/10.1111/emip.12359

Lilley, M., Meere, J., & Barker, T. (2016). Remote live invigilation: A pilot study. *Journal of Interactive Media in Education, 6*, 1–5. http://dx.doi.org/10.5334/jime.408

Liu, C., Han, K. T., & Li, J. (2019). Compromised item detection for computerized adaptive testing. *Frontiers in Psychology, 10*, 829. http://dx.doi.org/10.3389/fpsyg.2019.00829

Milone, A. S., Cortese, A. M., Balestrieri, R. L., & Pittenger, A. L. (2017). Theimpact of proctored online exams on the educational experience. *Currents in Pharmacy Teaching and Learning, 9*(1), 108–114. https://doi.org/10.1016/j.cptl.2016.08.037

Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020). Minimize online cheating for online assessments during COVID-19 pandemic. *Journal of Chemical Education, 97*(9), 3429-3435. https://doi.org/10.1021/acs.jchemed.0c00790

O'Shea, S., Stone, C., & Delahunty, J. (2015). "I 'feel' like I am at university even though I am online." Exploring how students narrate their engagement with higher education institutions in an

online learning environment. *Distance Education, 36*(1), 41–58. https://doi.org/10.1080/01587919.2015.1019970

Orr, R., & Foster, S. (2013). Increasing student success using online quizzing in introductory (majors) biology. *CBE—Life Sciences Education, 12*(3), 509–514. https://doi.org/10.1187/cbe.12-10-0183

Pagram, J., Cooper, M., Jin, H., & Campbell, A. (2018). Tales from the exam room: Trialing an e-exam system for computer education and design and technology students. *Education Sciences, 8*(4), 188. https://doi.org/10.3390/educsci8040188

Palloff, R. M., & Pratt, K. (2009). *Assessing the online learner: Resources and strategies for faculty*. John Wiley & Sons, Inc.

Prisacari, A. A., Holme, T. A., & Danielson, J. (2017). Comparing student performance using computer and paper-based tests: Results from two studies in general chemistry. *Journal of Chemical Education, 94*(12), 1822-1830. https://doi.org/10.1021/acs.jchemed.7b00274

Rahim, A. F. A. (2020). Guidelines for online assessment in emergency remote teaching during the COVID-19 pandemic. *Education in Medicine Journal, 12*(2), 59-68. https://doi.org/10.21315/eimj2020.12.2.6

Ravasco, G. G. (2012). Technology-aided cheating in open and distance e-learning. *Asian Journal of Distance Education, 10*(2), 71–77. https://www.learntechlib.org/p/185226/

Rios, J. A., & Liu, O. L. (2017) Online Proctored versus unproctored low-stakes internet test administration: Is there differential test-taking behavior and performance? *American Journal of Distance Education, 31*(4), 226-241, https://doi.org/10.1080/08923647.2017.1258628

Solak, H. İ., Ütebay, G., & Yalçın, B. (2020). Uzaktan eğitim öğrencilerinin basılı ve dijital ortamdaki sınav başarılarının karşılaştırılması. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi, 6*(1), 41-52. https://dergipark.org.tr/en/pub/auad/issue/55639/761293

Sorensen, E. (2013). Implementation and student perceptions of e-assessment in a Chemical Engineering module. *European Journal of Engineering Education, 38*(2), 172–185. https://doi.org/10.1080/03043797.2012.760533

Srikanth, M., & Asmatulu, R. (2014). Modern cheating techniques, their adverse effects on engineering education and preventions. *International Journal of Mechanical Engineering Education, 42*(2), 129-140. https://doi.org/10.7227/IJMEE.0005

Steger, D., Schroeders, U., & Gnambs, T. (2020). A meta-analysis of test scores in proctored and unproctored ability assessments. *European Journal of Psychological Assessment, 36*(1), 174–184. https://doi.org/10.1027/1015-5759/a000494

Sullivan, D. P. (2016). An integrated approach to preempt cheating on asynchronous, objective, online assessments in graduate business classes. *Online Learning, 20*(3), 195–209. https://doi.org/10.24059/olj.v20i3.650

Şenel, S, & Şenel, H. C. (2021). Remote assessment in higher education during COVID-19 pandemic. *International Journal of Assessment Tools in Education, 8*(2), 181-199. https://doi.org/10.21449/ijate.820140

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th Ed.). Pearson

Ullah, A., Xiao, H., & Barker, T. (2016) A classification of threats to remote online examinations. *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 1-7, https://doi.org/10.1109/IEMCON.2016.7746085

Ural, M. N., & Takaoğlu, Z. B. (2023). Comparison of online and face-to-face exams conducted in Physics I course in higher education. *Turkish Journal of Engineering*, *7*(1), 9-16. https://doi.org/10.31127/tuje.1004417

Weiner, J. A., & Hurtz, G. M. (2017). A comparative study of online remote proctoring versus onsite proctored high-stakes exams. *Journal of Applied Testing Technology, 18*(1), 13–20. https://jattjournal.com/index.php/atp/article/view/113061

Zacharis, N. Z. (2010). Innovative Assessment for Learning Enhancement: Issues and Practices. *Contemporary Issues in Education Research (CIER), 3*(1), 61–70. https://doi.org/10.19030/cier.v3i1.162

Zhang, Y., Wang, D., Gao, X., Cai, Y., & Tu, D. (2019). Development of a computerized adaptive testing for internet addiction. *Frontiers in psychology, 10*, 1010. https://doi.org/10.3389/fpsyg.2019.01010

About the Author(s)

- Necati Taşkın (Corresponding author); necatitaskin@odu.edu.tr; Ordu University; Türkiye; https://orcid.org/0000-0001-8519-6185
- Kerem Erzurumlu; keremerzurumlu@odu.edu.tr; Ordu University; Türkiye; https://orcid.org/0000-0001-5363-1963

Author's Contributions (CRediT)

Necati Taşkın: Statistical analysis, Visualization, Writing – original draft; Kerem Erzurumlu: Methodology, Data curation, Writing – review & editing.

Ethics Statement

The data used for this article are data registered in the learning management system. This study is carried out by the lecturers in the distance education application and research center of a state university. The task description of this institution includes application and research. No scale and/or questionnaire was applied to the students. The research does not violate ethical principles.

Conflict of Interest

The authors do not declare any conflict of interest.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Suggested citation:
Taşkın, N., & Erzurumlu, K. (2023). The Examination of Online and Paper-Pencil Test Scores of Students Engaged in Online Learning. *Asian Journal of Distance Education, 18*(2), 94-108. https://doi.org/10.5281/zenodo.8338263