



Effectiveness of Online Testing Versus Traditional Testing: A Comparative Study of Saudi Female College Students

Asma M. Abumalik^{a*}, Fatmah A. Alqahtani^b

^a *Department of Applied Linguistics, College of Languages,
Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.*
Email: amabumalik@pnu.edu.sa

^b *Department of Applied Linguistics, College of Languages,
Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.*
Email: faalghahtani@pnu.edu.sa

Received: 09 August 2023 | Received: in Revised form 11 November 2023 | Accepted 13 January 2024

APA Citation:

Abumalika, A. M., Iqahtanib, F. A. (2024). Effectiveness of Online Testing Versus Traditional Testing: A Comparative Study of Saudi Female College Students. *Eurasian Journal of Applied Linguistics*, 10(1), 151-159.
Doi: <http://dx.doi.org/10.32601/ejal.10114>

Abstract

Testing is an effective method to determine learning outcomes for knowledge and skills learning domains. The aim of this study was to examine the differences in test achievements among 50 Saudi female English major students at the College of Languages at Princess Nourah University in Riyadh. The tests were administered using two different methods: paper-based and Blackboard-based (online). Additionally, the study explored the impact of these two test methods on students' achievement in terms of course learning outcomes. The results of the study indicated that there was no significant difference between the two test methods in terms of overall test scores. However, it was found that the Blackboard-based test resulted in slightly higher scores for knowledge domain outcomes, while the paper-based test showed higher scores for skills domain outcomes. The results obtained in this study suggest that both paper-based and Blackboard-based test methods can be equally effective at assessing the general achievement of students. However, the choice of test method may have a slight impact on the specific learning outcomes being assessed, with Blackboard-based tests favouring knowledge domain outcomes and paper-based tests favouring skills domain outcomes. Furthermore, when using Blackboard-based tests, time pressure should be taken into consideration, as it is observed to significantly influence students' performance in both learning domains.

© 2024 EJAL & the Authors. Published by Eurasian Journal of Applied Linguistics (EJAL). This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Paper-Based Testing, Online-Based Testing, Blackboard, Traditional Testing, Learning Outcomes.

Introduction

In order to determine learning outcomes for knowledge and skills learning domains, testing is an essential component in the higher education system (Ahea, Ahea, & Rahman, 2016). This process, which involves providing effective feedback, is often seen as a challenging task. Traditionally, testing involves the use of paper, pen or pencil, where students answer questions on physical exam papers. However, with advancements in technology and the widespread availability of internet access, online testing has emerged as an alternative to the traditional paper-based approach (Shahini & Hashemi Toroujeni, 2023; Yao, 2020). Despite the internal

* Corresponding Author

Email: amabumalik@pnu.edu.sa

DOI: <http://dx.doi.org/10.32601/ejal.10114>

and external challenges presented by (Johnson, Jacovina, Russell, & Soto, 2016), which educators might encounter in integrating technology in the process of learning in general and of assessment in particular, this integration has become necessary. This necessity has resulted in the adoption of online testing as a viable alternative to traditional paper-based testing (PBT), which involves students using pens or pencils to write their answers on physical answer sheets. Online-based testing (OBT), on the other hand, refers to administering tests through digital platforms or computer-based systems, such as Blackboard, in the case of this study. This method allows students to complete exams using digital devices, such as computers, laptops, or tablets (Rogers et al., 2009). Both the OBT and PBT methods have their own set of advantages and disadvantages (Al Rawashdeh, Mohammed, Al Arab, Alara, & Al-Rawashdeh, 2021). While PBT methods provide a more controlled environment, OBT methods offer the convenience of being accessed from anywhere with an internet connection, which supports catering to the diverse needs of our students.

The great shift in the educational environment after COVID-19 in Saudi Arabia has encouraged researchers to conduct more research on online testing to raise its effectiveness of such tests. Among those studies was one conducted by Aljohani, Aloreafy, Alzaidi, and Meccawy (2021) on a sample of 49 Saudi female EFL students at the university of Jeddah. This study claimed that students' positive perception towards availability and accessibility, instructions, and mode of delivery of online testing could be raised by clear instructions and suitable mode of delivery. Alharbi, Ahmed, and Alhaj-Mahmoud (2021) believed that the faculty members need serious training on how to prepare an e-test, and efforts should be paid to prevent the occurrence of some defects including the starting time and the use of the system of delivering the test that might negatively affect the performance of students whether males or females.

In Saudi Arabia, there is a dearth of studies that have investigated the factors that might affect testing, the core of the process of teaching. Nor has been any study that could examine how an improvement in testing can lead to improvement in the measurements of the learning outcomes targeted by each course. The current study, therefore, aimed to fill this research gap, and addressed this educational concern in a comparative fashion. The aim of the present study was to investigate the potential differences in students' performance on a paper test versus an online test designed on the Blackboard Learning Management System (LMS).

This study compared the effectiveness of OBT and PBT in assessing students' level of achievement in terms of course learning outcomes in knowledge and skills domains. Factors such as marks and completion time and their impact on students' performance were examined, which provided insights for educators in implementing both methods effectively. Additionally, the potential implications of the final findings of the present study were discussed. By analysing the strengths and weaknesses of both online and traditional testing methods concerning the main factors of the study, educators can make informed decisions regarding which approach best suits their own needs and students' needs. The data collected from this study would provide valuable insights into the potential advantages and disadvantages of using online testing platforms, such as Blackboard, in educational settings.

Literature Review

The Occurrence of Online Testing in Language Teaching

The history of online language testing can be traced back to the late 1990s and early 2000s when the internet began gaining popularity and accessibility (Kunnan, 2017). Prior to this, language testing was primarily conducted in traditional classroom settings or through paper-based exams. In the early stages, online language testing focused mainly on multiple-choice questions and automated scoring systems. These tests were primarily used for assessing vocabulary, grammar, and reading comprehension skills; however, they were limited in their ability to evaluate speaking and writing abilities, which required more subjective evaluation (Fulcher & Davidson, 2007). Over time, advancements in technology and the increasing demand for comprehensive language assessment led to the development of more sophisticated online testing platforms. These platforms incorporated features such as audio and video recordings, interactive speaking and writing tasks, and real-time feedback (McCraw, 2017).

In 2020, universities worldwide were compelled to swiftly transition to complete online education due to the COVID-19 pandemic. This has posed a significant hurdle in delivering lectures, seminars and other forms of teaching that were traditionally conducted in person. Additionally, the issue of conducting online assessments emerged as a concern (Selwyn, O'Neill, Smith, Andrejevic, & Gu, 2023). The emergence of online language testing was driven by the need for a more convenient and efficient way to assess language proficiency since this process is considered the core of any education system (Aravind, Bhuvanewari, & Rajest, 2023). Although the debate over the advantages and disadvantages of online testing continues, online testing offers several advantages, such as flexibility in scheduling, accessibility from anywhere with an internet connection, and immediate feedback (Singh, 2021). These advantages make this method of assessment a strong option for educators.

Today, online language testing has become widely accepted and recognized as a valid and reliable method of assessing language proficiency; it is used by educational institutions, language learning centers, and

employers to evaluate language skills. Additionally, conducting such tests has become much easier due to the widespread availability of internet connections and the familiarity of students with technology. The purpose of this research was not to persuade educators to adopt online testing but rather to guide them toward the most effective practices and strategies for utilizing online testing to assess students' achievement of the desired learning outcomes.

Online Testing vs. Traditional Testing

Online education and testing have gained immense popularity in recent years, prompting numerous research studies comparing computer-based tests (CBTs) with paper-based tests (PBTs) in the field of language assessment (Yusupjonova & Axmadaliyeva, 2023). With the rapid advancement of technology and the ubiquitous availability of internet access, educational institutions have increasingly embraced the use of online learning platforms. These platforms have been extensively evaluated to gauge their effectiveness and advantages over traditional classroom-based learning (Shahini & Hashemi Toroujeni, 2023; Yao, 2020). Over the last few decades, multiple research investigations have assessed various aspects, such as student performance, engagement, satisfaction, and overall learning experience, aiming to enhance teaching and assessment processes. In this context, the current study offers a comprehensive review of the literature focusing on comparisons between traditional testing methods and online testing methods.

Chua (2012) conducted a study to investigate the effectiveness of CBT on test performance, testing time and testing attitude by comparing it with that of PBT. The results showed that the CBT technique is more reliable in terms of internal and external validity and significantly reduces testing time; additionally, it creates more self-efficacy and a positive attitude towards the testing process. Since online testing has largely replaced traditional testing for some courses in the Department of Informatics at the University of Rejika, a study by Candrljic, Katić, and Dlab (2014) conducted a comparative study on three of those courses to evaluate the differences between results on paper-based testing versus online testing using MudRi. Their study included 1231 paper-based and online tests. When a certain percentage of questions were subjective, the results of the study showed no significant difference between the MudRi online test and traditional paper-based test results.

Öz and Özturan (2018) investigated whether computer-based vs. paper-based test-delivery modes had an impact on the reliability and validity of an achievement test for a pedagogical content knowledge course in an English teacher education program. A total of 97 university students enrolled in the English as a foreign language (EFL) teacher education program were randomly assigned to the experimental group that took the computer-based achievement test online and the control group that took the same test in a paper-and-pencil format. The results indicated that the test delivery mode did not have any impact on the reliability or validity of the tests administered in either way. The findings also demonstrated that there was no significant difference in test scores between participants who took the computer-based test and those who took the paper-based test.

In a study conducted by E AlAdl (2020), the impact of utilizing e-tests on the performance of students in a preparatory English course at Delta University was examined. Additionally, the study recorded the attitudes of the students towards the e test. The 60 participants were divided into two groups: an experimental group evaluated through e test methods and a control group evaluated through traditional paper-based tests. The results indicated that e-tests had a more positive effect on students' performance than paper-based tests. Furthermore, participants displayed a positive attitude towards e-tests, likely due to their familiarity with various technological platforms.

In a study conducted by Yu and Iwashita (2021), 92 Chinese undergraduate English major students were randomly assigned to either the CBT or PBT group. To gather the data, a mixed methods approach was employed, which involved both quantitative and qualitative analysis of test performance in the two modes, as well as an assessment of the CBT group participants' computer familiarity and attitudes towards the mode. Additionally, semi-structured interviews were conducted and subjected to thematic analysis. The findings of the study indicated that there was no significant difference in test scores between the CBT and PBT groups. Furthermore, two items from the computer familiarity questionnaire, namely, the comfort level when reading articles on the computer and the amount of time spent using computers, were found to be positively correlated with CBT scores. Interestingly, the participants' attitudes towards CBT did not appear to have an impact on their test performance.

Cai (2022) performed a study with regard to participants' experience and personal opinions on online assessments and the impact of online assessments on their grades. Both quantitative and qualitative research were adopted. Three tests and a questionnaire survey were used for evaluating the impact of online assessments on the performance of the participants, and data on their views on different online assessment formats were collected. Qualitative research was applied to interview the respondents to obtain additional details on their opinions on the online assessments. The findings of the study indicated that online assessments have a negative impact on performance and that candidates prefer traditional testing on paper.

Although the study conducted by Kucherova and Ushakova (2022) involved assessing attitudes toward

online language testing during the COVID-19 pandemic, the findings can also be utilized by university teachers in blended or face-to-face teaching. They aimed to analyse the effectiveness of online language testing and compare the attitudes of students and teachers towards online testing in a general English university course. Online tests were administered as a synchronous component of distance learning to 857 first-year bachelor's degree students by 20 teachers during the 2020-2021 academic years. The authors collected data through an online questionnaire completed by students and teachers. The results showed that different types of questions could be successfully applied to online language testing and that online language testing can be effective and relevant to course objectives.

Studies on Online Testing in Saudi Arabia

Al-Qdah and Ababneh (2017) investigated the effects of online testing and paper testing on male and female college students. Tests were designed with various types of questions: multiple-choice questions (MCQs), true/false (T/F), short and essay type questions. The statistical results found no significant difference between the two tests concerning the objective questions, MCQs and T/F, but when it comes to writing, students prefer to write on paper rather than on screen. In her study, Alghamdi (2022) pointed to the effect of test media on the level of test anxiety. She used a questionnaire to measure the level or percentage of students' test anxiety. The results were on the side of online tests. The test anxiety was higher when taking a test on paper rather than online.

Since the implementation of E-tests in general has become a necessity in higher education as Gokulkumari, Gokulkumari, Al-Hussain, Akmal, and Singh (2022) stated, they felt the importance of measuring students' perceptions towards E-Exam system in higher education in the universities of Saudi Arabia. The need for quick and fast results, continuous measurement of students' progress and variety in delivering assessments were behind conducting this study. The tool of collecting data which was a questionnaire was distributed among students and researchers in Saudi universities. The study aimed to provide valuable information about E-test concerning its implementation, environment, students' previous experience and the challenges facing its systems. Based on the participants' personal experiences, recommendations for E-test requirements were proposed.

In another study, the factor of test anxiety was tackled, conducted on 826 students enrolled in different colleges at King Faisal University in Saudi Arabia. This study emphasized the necessity of having mobile exam platforms (MEPs). The answers of randomly chosen sample on the survey revealed that the majority preferred MEPs for reasons that included Ease to access and less test anxiety (Salem & Alshebami, 2023).

The studies reviewed in this literature significantly contribute to our understanding of online testing vs. traditional testing and offer valuable insights for educators and researchers in the field. By exploring the factors that influence performance in testing, these studies shed light on how different modes of assessment may affect students' learning outcomes. However, there is a dearth of studies that would examine students' achievement and learning outcomes in a comparative way in the Saudi context. This gap in literature is related to comparison between paper based and online testing methods. The current study aimed to bridge this gap. The findings of this study would be crucial for educators as it would enable them to design effective testing strategies and optimize learning experiences. Furthermore, researchers can use these findings to further investigate the dynamics between testing modes and performance.

Based on the literature survey and the compelling need to fill the research gap, this study framed the following hypotheses to achieve the study objectives.

H1: *Student Performance on Paper is Better than that on Blackboard.*

H2: *Student Achievement in Terms of Course Learning Outcomes is Greater when Assessed through Paper-Based Tests than when Assessed through Blackboard-Based tests.*

H3: *The Time Pressure is Significantly Greater when Tested on Blackboard.*

Methodology

Research Design

The quantitative research design was used in this study with a structured approach to compare students' marks on both online and paper-based tests and determine if there was a significant difference. This method has been used in conjunction with qualitative methods to gain a comprehensive understanding of the factors that might have influenced the students' performance on the two tests. The performance of the students on both tests was assessed based on their grades, completion times, and potential discrepancies in their responses.

Sampling

The sample of the present study consisted of 50 undergraduate students at the first level of college. All participants were English as foreign language (EFL) majors at the College of Languages at Princess Nourah University. All participants were Saudi females aged between 18 and 20 years. The participants were

assigned to two groups: one group took a traditional paper-based test, while the other group took an equivalent test administered online through the Blackboard platform.

Instruments of the Study

To achieve the aims of the study, the researchers used specific instruments including two tests, observation methods and Group discussion. The two tests were performed: one on paper and the other on the Blackboard platform. Both tests consisted of an equivalent set of objective questions and covered the same content. The questions were multiple-choice and true/false. The observation method was used to effectively record anything that was worth noting in affecting the students' performance on the two tests. The Group Discussion was a platform where were discussed the results of the two forms of testing with the students to help them understand the reasons behind any noticeable differences in marks and to provide a deeper understanding of the factors that might have influenced their performance.

Data Analysis

The t test was used for data analysis since it fit the aim of the present study. The test was used to compare the two variables of the study (testing on paper and testing on Blackboard) for the assumption of having any statistically significant differences between the means of the two groups.

Results

Both the tests, online test (Blackboard) and paper based, were administered with the same content with both groups. Each test was divided into three parts: Part I (5 points), Part II (15 points) and Part III (10 points). The three parts measured different course learning outcomes. The participants in the Blackboard group took the exam online, and the test results were automatically corrected by the system. The paper group, on the other hand, had their tests manually corrected by the researchers. The researchers recorded the results on the three parts. Finally, the total scores for all the students were calculated as shown in [Table 1](#).

Table 1: Students' Scores on the Two Tests.

Online test (Blackboard)					Paper based Test				
Student ID	Part I	Part II	Part III	Total	Student ID	Part I	Part II	Part III	Total
445008004	4	14	8	26	445009623	3	8.5	6	17.5
445007967	4	10	8	22	445007997	4	13	9	26
445007961	4	11	6	21	445009380	4	8	6.5	18.5
445007976	4	10	10	24	445008024	4	12	5.5	21.5
445008032	3	10	8	21	445008025	4	10	6.5	20.5
445007994	4	9.5	6	19.5	445007990	5	12.5	9	26.5
445007968	5	12	8	25	445008002	4	11.5	7.5	23
445007977	3	11.5	9	23.5	445007974	4	10	5	19
445008016	3	8.5	4.5	16	445007988	5	14	9	28
445007955	4	13.5	9	26.5	445007966	5	9.5	7	21.5
445008035	3	10	6	19	445008003	4	10.5	7.5	22
445008019	3	8	8	19	445007982	5	12	9	26
445007965	5	15	10	30	445007964	5	11.5	9	25.5
445008026	4	10.5	8	22.5	445007953	4	13.5	10	27.5
445007954	4	11.5	8	23.5	445007980	4	10.5	8.5	23
445007996	4	8	9	21	445007992	5	11	9	25
445008013	4	10.5	6	20.5	445009534	4	14	10	28
445007948	3	9.5	6	18.5	445007960	4	13	9	26
445007999	4	8.5	8.5	21	445008020	1	13.5	8.5	23
445008012	5	9	9	23	445008006	5	13	8.5	26.5
443004694	4	13	10	27	445007985	5	11.5	7	23.5
445008022	5	10	7	22	445008005	4	6.5	7	17.5
445009375	5	9	10	24	445008030	4	12.5	6	22.5
445007973	4	12.5	8	24.5	445007993	5	15	9	29
445010037	4	10.5	5.5	20	445008014	3	8.5	6	17.5

The aim was to identify any significant differences between the two media in terms of test scores. The average total score of students in the Blackboard group was compared to that of students in the paper group to determine whether there was a significant difference in performance between the two groups. Additionally, the scores obtained by students in each of the three parts, corresponding to the learning outcomes measured by the test, were compared between the Blackboard and paper groups. This comparison was intended to statistically identify any discrepancies in performance based on specific learning outcomes.

In general, and based on the T test results, there was a slight difference between the means of testing on

paper (23.38) and testing on Blackboard (22.4), as shown in [Table 2](#). In other words, the analysis showed that the performance of students who had the test on paper was slightly better than that of those who had the test on Blackboard. However, at the 0.05 level, this difference was not statistically significant.

Table 2: *T Test Analysis of the Students' General Performance.*

	Variable 1= paper	Variable 2= Blackboard
Mean	23.38	22.4
Variance	12.65166667	9.708333333
Observations	25	25
Pearson Correlation	-0.173129942	
Hypothesized Mean Difference	0	
Df	24	
t Stat	0.957339891	
P(T<=t) one-tail	0.173970425	
t Critical one-tail	1.71088208	

Concerning the three different parts of the test, several differences between the two groups were observed. For instance, Part I measures the students' performance in the test on the knowledge domain. [Table 3](#) shows the statistical analysis of these data.

Table 3: *T Test Analysis of the Students' Performance on Part I.*

	Variable 1= paper	Variable 2= Blackboard
Mean	4.16	3.96
Variance	0.806666667	0.456666667
Observations	25	25
Pearson Correlation	0.010984038	
Hypothesized Mean Difference	0	
Df	24	
t Stat	0.894427191	
P(T<=t) one-tail	0.189987956	
t Critical one-tail	1.71088208	

According to the results of the t test, the P value is 0.189987956, and the T value is 1.71088. The P value represents the probability of obtaining the observed difference in means if the null hypothesis is true. In this case, since the P value is greater than the commonly used significance level of 0.05, there is not enough evidence to conclude that there is a significant difference between the two tests (paper and Blackboard). The t- value, on the other hand, represents the magnitude of the difference between the means relative to the variability within the groups. A higher t- value indicates a larger difference between the means. In this case, the t- value of 1.71088 suggested a moderate difference between the means of the two tests. In summary, based on the T test results, there was no significant difference between the paper and Blackboard tests, as indicated by the P value of 0.189987956. However, there was a moderate difference between the means, as suggested by the t- value of 1.71088.

Part II part measures students' performance in the test on the skills domain. [Table 4](#) reveals the statistical analysis of these data.

Table 4: *T Test Analysis of the Students' Performance on Part II.*

	Variable 1= paper	Variable 2= Blackboard
Mean	11.42	10.62
Variance	4.493333333	3.505833333
Observations	25	25
Pearson Correlation	-0.273054387	
Hypothesized Mean Difference	0	
Df	24	
t Stat	1.254500096	
P(T<=t) one-tail	0.110870337	
t Critical one-tail	1.71088208	

The p value is 0.110870337, and the t value is 1.71088,208, which means that the p value is greater than the significance level of 0.05. A p- value greater than the significance level suggests that there is not enough evidence to reject the null hypothesis. Therefore, based on these results, there is insufficient evidence to conclude that there is a significant difference or relationship between the variables being tested.

Part III also measures students' performance in the test on the skills domain, testing a different set of skills. [Table 5](#) represents the statistical analysis of these data.

Table 5: T-Test Analysis of the Students' Performance on Part III.

	Variable 1= paper	Variable 2= Blackboard
Mean	7.8	7.82
Variance	2.125	2.455833333
Observations	25	25
Pearson Correlation	-0.248967408	
Hypothesized Mean Difference	0	
Df	24	
t Stat	-0.041818193	
P(T<=t) one-tail	0.483494796	
t Critical one-tail	1.71088208	

The mean for the paper test is seen 7.8, while the mean for the Blackboard test is 7.82. The t value obtained from the test is 1.71088, which indicates a slight difference between the two groups, with the Blackboard group having a slightly greater mean score. The p value associated with the t test is calculated to be 0.483494796, which is greater than the commonly used significance level of 0.05; thus, the null hypothesis was rejected. This means that there is not enough evidence to conclude that the means of the two variables are significantly different. In summary, based on the t test results, there is a slight difference between testing on paper and testing on Blackboard, but this difference is not statistically significant.

Based on the above analysis, the hypotheses of the study were addressed:

H1: Students' Test Performance on Paper is Better than that on Blackboard.

The comparison between students' performance on paper and their performance on a Blackboard showed that their overall performance was not statistically significant. However, when testing on paper, it was found that students achieved slightly better results; this implies that when students were assessed using written exams or assignments, they performed slightly better than when they were assessed using Blackboard. Hence, this hypothesis was accepted.

H2: Students' Achievements in Terms of Course Learning Outcomes are Greater when Assessed through Paper-Based Tests Than when Assessed through Blackboard-Based Tests.

The results of the study indicated that there was a slight difference in the mean scores of the learning outcomes between the two groups being compared. Specifically, regarding knowledge, there was no statistically significant difference between the two groups. However, it was observed that the group that took the test on paper had a moderately greater score than the other group; when examining the skill outcomes, the statistical analysis revealed that there was no significant difference between the means of the two groups; this means that the performance of both groups in terms of skills was relatively similar. However, it is worth noting that the group that used Blackboard had a slightly greater mean score than did the other group. This suggests that this hypothesis had a mixed interpretation. While there was no statistically significant difference in learning outcomes between the two groups, there were some subtle variations. The paper test group performed slightly better in terms of knowledge, while the Blackboard group had a slightly greater mean score for skill outcomes.

H3: The Time Pressure is Significantly Greater when Tested on Blackboard.

During the test on Blackboard, unlike when tested on paper, the time pressure is felt significantly greater. Even though the students were familiar with the Blackboard platform and had no registered technical issues as they were performing their test, the pressure of time was a significant factor. Students, as they reported, had to deal with unexpected interruptions and delays; this might result from the lack of trust in technology. Additionally, they said that they feared technical glitches and connectivity issues. Moreover, the existence of a physical clock on the screen worsened the situation. Hence, this hypothesis was accepted.

Discussion

This study revealed a slightly better performance of students on paper in terms of knowledge outcomes indicating that paper tests might enhance students' power of retention of information greater than the Blackboard test. The Blackboard test which although demonstrated better practical application potential influence, it was found that this influence was limited to only specific skills or knowledge domains. This study also suggested that the difference in performance between paper test and Blackboard test is not statistically significant. There is only a slight difference observed on paper test, which suggests that the paper mode of assessment may have some influence on students' achievement. These findings are however contrary to the findings of [Kucherova and Ushakova \(2022\)](#), which found online language testing more effective and relevant in the attainment of learning outcomes. Further analysis or investigation however may be needed to understand the factors contributing to this difference and its implications for teaching and learning.

The study also reveals that the pressure of time proves to be a challenge for online tests. The lack of trust in technical tools and fear of unexpected interruptions and delays are the reasons reported in the present

study. Having these worries, teachers are supposed to try to eliminate any other distractions or external factors that might add to the sense of students' time pressure. Therefore, it is worth considering the overall test environment in which the students are taking the test. It is important for teachers to consider providing appropriate support and guidance to help their students manage their time effectively during assessments. Chua (2012), too, agreed on the time requirement for predicting a good performance. The study showed that having enough time in the tests ensures internal and external validity and creates more self-efficacy and positive attitude towards the testing process.

These findings support the idea of incorporating variety in testing methods to fully capture different aspects of student learning. By using a combination of paper-based and Blackboard-based tests, educators can maximize the strengths of each method, ensuring a comprehensive assessment of students' knowledge and skills. Ultimately, this promotes a more comprehensive and balanced evaluation of students' knowledge and skills. The results indicated that the test delivery mode did not have any impact on the reliability or validity of the tests administered in either way. The findings also demonstrated that there was no significant difference in test scores between participants who took the computer-based test and those who took the paper-based test. In practical terms, this means that educators can use paper-based tests, such as problem-solving or hands-on tasks, to assess students' skill-based performance. Blackboard-based tests can also be utilized for evaluating students' theoretical knowledge or for memorization-based assessments. This is consistent with the study by Oz and Turan (2018), who discovered that whether computer-based or paper-based test, there is no impact on the reliability and validity. However, Cai (2022) differed in the opinion and believed that online assessments rather have a negative impact on performance as compared to traditional paper testing method.

Conclusion

The aim of the present study was to compare students' performance on tests administered on paper and on Blackboard. A statistical analysis of the data revealed no significant difference in students' general performance between the two methods. This finding suggested that both methods are equally effective at assessing overall performance. Furthermore, the study investigated a slight difference in measuring course learning outcomes, specifically knowledge and skills. Although no statistically significant difference was found, it was observed that testing knowledge was slightly more effective on Blackboard.

The significance of the present study lies in its investigation of the effect of online testing versus paper-based tests on the performance of Saudi female college-level students. With the increasing integration of technology in education, it is crucial to understand how different modes of assessment impact students' performance on tests. By comparing the two testing methods, researchers can gain insights into the potential benefits and drawbacks of both methods of assessment, which can guide educators in making decisions about suitable assessment practices.

Furthermore, this study contributes to the existing body of research on online testing. By analysing the performance of students who took online tests versus those who took paper-based tests, the researchers can determine whether the mode of assessment influences students' ability to demonstrate their knowledge and skills accurately. This information is valuable for educators who are considering implementing online testing as part of their assessment strategies, as it provides evidence-based insights into its potential impact on students' achievements.

By adopting a varied approach to testing, educators can enhance the accuracy and reliability of their assessments, providing a more comprehensive understanding of students' learning outcomes; this, in turn, contributes to more effective teaching and supports students' overall academic development.

References

- Ahea, M. M.-A.-B., Ahea, M. R. K., & Rahman, I. (2016). The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education. *Journal of Education and Practice*, 7(16), 38-41. doi: Retrieved from <https://core.ac.uk/download/pdf/234639018.pdf>
- Al-Qdah, M., & Ababneh, I. (2017). Comparing online and paper exams: Performances and perceptions of Saudi students. *International Journal of Information and Education Technology*, 7(2), 106-109. doi: <https://doi.org/10.18178/ijiet.2017.7.2.850>
- Al Rawashdeh, A. Z., Mohammed, E. Y., Al Arab, A. R., Alara, M., & Al-Rawashdeh, B. (2021). Advantages and disadvantages of using e-learning in university education: Analyzing students' perspectives. *Electronic Journal of E-learning*, 19(3), 107-117. doi: <https://doi.org/10.34190/ejel.19.3.2168>
- Alghamdi, N. A. (2022). Saudi Undergraduate Students' Preferences toward Tests and Test Anxiety Differences in Blended Learning. *International Journal of Linguistics, Literature and Translation*, 5(4), 107-115. doi: <https://doi.org/10.32996/ijllt.2022.5.4.13>

- Alharbi, I., Ahmed, S. A., & Alhaj-Mahmoud, F. M. (2021). Students' Attitudes at the University of Jeddah Towards E-Tests. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 12(13), 1-16. doi: <http://doi.org/10.14456/ITJEMAST.2021.268>
- Aljohani, A., Aloreafy, M., Alzaidi, S., & Meccawy, Z. (2021). Saudi EFL students' perceptions of online English achievement exams in the era of COVID-19. *International Journal of Linguistics, Literature and Translation*, 4(5), 118-125. doi: <https://doi.org/10.32996/ijllt.2021.4.5.11>
- Aravind, B., Bhuvanewari, G., & Rajest, S. S. (2023). ICT-based digital technology for testing and evaluation of English language teaching. In *Handbook of Research on Learning in Language Classrooms Through ICT-Based Digital Technology* (pp. 1-11): IGI Global. doi: <https://doi.org/10.4018/978-1-6684-6682-7.ch001>
- Cai, Y. (2022). A study on the implementation of online assessments in college English teaching. In *SHS Web of Conferences* (Vol. 140, pp. 01006): EDP Sciences. doi: <https://doi.org/10.1051/shsconf/202214001006>
- Candrljic, S., Katić, M. A., & Dlab, M. H. (2014). Online vs. Paper-based testing: A comparison of test results. In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 657-662): IEEE. doi: <https://doi.org/10.1109/MIPRO.2014.6859649>
- Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, 28(5), 1580-1586. doi: <https://doi.org/10.1016/j.chb.2012.03.020>
- E AlAdl, A. (2020). Using electronic tests versus pen and paper tests: the experience of Delta University. *Journal Of the Faculty of Education- Mansoura University*, 110(2), 31-45. doi: <https://dx.doi.org/10.21608/maed.2020.147690>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. Routledge London and New York. doi: Retrieved from <https://www.routledge.com/Language-Testing-and-Assessment-An-Advanced-Resource-Book/Fulcher-Davidson/p/book/9780415339476>
- Gokulkumari, G., Al-Hussain, T., Akmal, S., & Singh, P. (2022). Analysis of E-Exam practices in higher education institutions of KSA: Learners' perspectives. *Advances in Engineering Software*, 173, 103195. doi: <https://doi.org/10.1016/j.advengsoft.2022.103195>
- Johnson, A. M., Jacovina, M. E., Russell, D. G., & Soto, C. M. (2016). Challenges and solutions when using technologies in the classroom. In *Adaptive educational technologies for literacy instruction* (pp. 13-30): Routledge. doi: <https://doi.org/10.4324/9781315647500-2>
- Kucherova, O., & Ushakova, I. (2022). Effectiveness of online testing in General English University Course from teacher and student perspectives. *Information Technologies and Learning Tools*, 87(1), 185-198. doi: <https://doi.org/10.33407/itlt.v87i1.4812>
- Kunnan, A. J. (2017). *Evaluating Language Assessments*. Taylor & Francis. doi: <https://doi.org/10.4324/9780203803554>
- McCraw, D. B. (2017). *A Brief History of Testing: We've Come a Long Way!* CompTIA. doi: Retrieved from <https://www.comptia.org/blog/a-brief-history-of-testing-we-ve-come-a-long-way!>
- Öz, H., & Özturan, T. (2018). Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies*, 14(1), 67-85. doi: Retrieved from <https://www.researchgate.net/profile/Huseyin-Oz/publication/324112270>
- Rogers, P. L., Berg, G. A., Boettcher, J. V., Howard, C., Justice, L., & Schenk, K. D. (2009). *Encyclopedia of distance learning*. IGI Global. doi: Retrieved from <https://www.igi-global.com/book/encyclopedia-distance-learning-second-edition/352>
- Salem, M. A., & Alshebami, A. S. (2023). Exploring the Impact of Mobile Exams on Saudi Arabian Students: Unveiling Anxiety and Behavioural Changes across Majors and Gender. *Sustainability*, 15(17), 12868. doi: <https://doi.org/10.3390/su151712868>
- Selwyn, N., O'Neill, C., Smith, G., Andrejevic, M., & Gu, X. (2023). A necessary evil? The rise of online exam proctoring in Australian universities. *Media International Australia*, 186(1), 149-164. doi: <https://doi.org/10.1177/1329878X211005862>
- Shahini, G., & Hashemi Toroujeni, S. M. (2023). Computer-Based Testing vs. Paper-Based Testing: Score Equivalence and Testing Administration Mode Preference. *Iranian Journal of English for Academic Purposes*, 12(3), 17-35. doi: Retrieved from https://journalscmu.sinaweb.net/article_181854.html
- Singh, M. (2021). The Advantages and Disadvantages of an Online Examination System. Retrieved from <https://blog.mettl.com/advantages-disadvantages-online-examination/>
- Yao, D. (2020). A Comparative Study of Test Takers' Performance on Computer-Based Test and Paper-Based Test across Different CEFR Levels. *English Language Teaching*, 13(1), 124-133. doi: Retrieved from <https://ideas.repec.org/a/ibn/eltjnl/v13y2020i1p124.html>
- Yu, W., & Iwashita, N. (2021). Comparison of test performance on paper-based testing (PBT) and computer-based testing (CBT) by English-majored undergraduate students in China. *Language Testing in Asia*, 11(1), 32. doi: <https://doi.org/10.1186/s40468-021-00147-0>
- Yusupjonova, D., & Axmadaliyeva, D. (2023). Using computer and online technologies in teaching english. *Models and methods in modern science*, 2(12), 151-155. doi: Retrieved from <http://econferences.ru/index.php/mmms/article/view/9983>