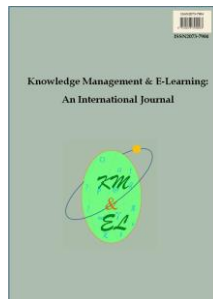

The impact of computer-based testing on EFL university students' reading and listening skills test performance and test anxiety levels

Andrew Ghaly

Pharos University in Alexandria, Egypt
Alexandria University, Egypt



Knowledge Management & E-Learning: An International Journal (KM&EL)
ISSN 2073-7904

Recommended citation:

Ghaly, A. (2026). The impact of computer-based testing on EFL university students' reading and listening skills test performance and test anxiety levels. *Knowledge Management & E-Learning*, 18(3), 493–508.
<https://doi.org/10.34105/j.kmel.2026.18.022>

The impact of computer-based testing on EFL university students' reading and listening skills test performance and test anxiety levels

Andrew Ghaly* 

Pharos University in Alexandria, Egypt
Faculty of Education
Alexandria University, Egypt
E-mail: andrew.wagdy@pua.edu.eg

*Corresponding author

Abstract: This study investigated the impact of computer-based testing (CBT) on upper-intermediate EFL university students' test performance in reading and listening, and on their test anxiety levels. Forty students at the English Language Centre of Pharos University in Alexandria ($N = 40$) were randomly assigned to a control group taking paper-and-pencil tests or an experimental group taking the same tests in a computer-based format. Three reading and three listening tests, each developed according to Cambridge B1 specifications, were administered over eight weeks. Students' test anxiety was assessed using the Test Anxiety Inventory questionnaire and semi-structured interviews. The results revealed no statistically significant differences in reading or listening performance between groups; however, test anxiety levels were significantly higher among CBT participants ($t = -8.99, p < 0.001$). The findings suggest that while CBT can be used interchangeably with traditional paper-based tests, additional measures such as pre-test orientation and digital familiarization sessions are essential to mitigate increased anxiety from CBT. For practitioners, this study highlights the importance of preparing and orienting EFL learners before transitioning to computer-based assessment environments.

Keywords: Computer-based testing; Paper-and-pencil testing; Language assessment; Anxiety; University students

Biographical notes: Andrew Ghaly is an educational manager, assessment specialist, and teacher trainer based in Alexandria, Egypt. He holds a Master's degree in Education (English Language Assessment) and has extensive experience in language assessment design, item writing, test validation, and computer-based testing. He works with international assessment organizations such as Trinity College London and Cambridge Assessment English and serves as Administrative Director of the English Language Centre at Pharos University in Alexandria where he established the centre's testing unit. His research and professional interests focus on assessment literacy, data-informed educational improvement, digital transformation, and AI-informed assessment and teacher training. He is a doctoral candidate of the Faculty of Education, Alexandria University.

1. Introduction

The ongoing trend in implementing and utilizing technology in teaching was entailed by a similar movement in the field of assessment. This desire among educators and researchers to integrate technology in education rooted from the tendency towards a more flexible, standardized, individualized learning and testing experience thanks to recent technological advances used in developing assessment tools. Those tools use computers and modern technology to deliver and score tests and in some cases, report scores and interpret them (Epstein & Klinkenberg, 2001; Ulfa et al., 2025; Wannas & AbdelMohsen, 2025). Additionally, the inclination to reduce costs and paper usage in educational institutions by phasing out test booklets and answer sheets underscores the high potential of computerised tests to help in the spread of green initiatives by universities and schools.

In recent years, Egypt has taken some serious steps towards the integration of assessment practices in the educational process through the implementation of formative assessment in lower grades and changing the assessment system in the general certificate of secondary education exams. In order to support student development and produce positive outcomes from the current improvement efforts, the Strategic Plan of Pre-University Education 2014–2030 (Ministry of Education, 2014) aims to modernise and reform the entire system of assessment and establishes contemporary evaluation systems.

Those policies include providing all secondary stage students with tablets through which they will receive their assessment in a mobile or computer-based mode. This sheds light on how important it is to study the effects of these changes in the major areas of test delivery. Besides, the Egyptian Ministry of Higher Education and Scientific Research has started implementing policies supporting universities to develop digital infrastructure and raise internet speeds in governmental universities. This contributes to their transformation to smart universities, in implementation of the general strategies of the ministry to maximize the use of computer-based automated tests for all university students in all specializations (Hu & Cao, 2025). Despite these national initiatives, empirical research investigating how Egyptian EFL learners respond to the shift from paper-and-pencil testing (PPT) to CBT remains scarce.

International studies present mixed findings: while some report comparable performance across the two modes (Al-Amri, 2008; Boo, 1997; Bugbee, 1996; Whitworth, 2001), others reveal the opposite (Pomplun et al., 2002; Choi et al., 2003; Scheuermann & Björnsson, 2009). However, little is known about whether these findings generalize to the Egyptian context, where students' familiarity with computers, institutional readiness, and cultural attitudes toward testing differ markedly from those in high-tech educational environments.

Therefore, the present study addresses this research gap by investigating whether the mode of test administration (CBT vs. PPT) affects Egyptian EFL university students' performance in reading and listening, and whether it influences their levels of test anxiety. This study also seeks to contribute locally relevant evidence that can inform the implementation of technology-driven assessment reforms in Egypt's higher education system.

The study is guided by the following research questions:

1. What is the impact of computer-based language testing on upper-intermediate university students' performance in reading comprehension?
2. What is the impact of computer-based language testing on upper-intermediate university students' performance in listening comprehension?

3. What is the impact of computer-based language testing on upper-intermediate university students' anxiety levels?

2. Literature review

2.1. Paper-and-pencil tests vs. computer-based tests

The transition from Paper-and-Pencil Tests (PPTs) to Computer-Based Tests (CBTs) requires establishing score comparability and construct equivalence. Many researchers assert that before CBTs can replace PPTs, the computerized versions must demonstrate equivalent levels of validity and reliability (Johnson & Green, 2006). Consequently, numerous studies have called for further check regarding comparability (Wise & Plake, 1990; Bridgeman et al., 2001; Parshall et al., 2002; Pomplun & Custer, 2005; Kingston, 2008).

However, establishing true equivalence is complicated due to several mode effects that threaten the consistency of the construct being measured. The comparability of scores, particularly for constructed-response items, may be affected by the input devices used and the examinees' computer literacy or familiarity (Alderson, 2000). For instance, a student's typing speed could impact performance to the extent of making the CBT measure a different construct compared to the PPT (Sandene et al., 2005; Kingston, 2008).

The socioeconomic effect also makes comparability harder, as differences in access to technology can lead to significant differences in students' performance and familiarity with the testing mode (MacCann, 2006). This threatens test standardization when institutions lack preparedness or infrastructure (Kingston, 2008).

Furthermore, factors related to the display, such as limited screen space (compared to a printed page), screen size, font type, and resolution, may distort the task so fundamentally that the CBT and PPT modes begin to measure different constructs entirely (McKee & Levinson, 1990; Bridgeman et al., 2001; Pommerich, 2004).

While the literature generally acknowledges the potential for mode effects (Leeson, 2006), research often attempts to attribute these effects only to procedural or technological incompatibility (Mayes et al., 2001; Noyes & Garland, 2003). A significant gap exists in studies that critically evaluate how the mode changes the cognitive processing in specific skills, like reading and listening comprehension.

For example, in listening comprehension, the contemporary paradigm argues for incorporating rich contextual elements (Buck, 2001), yet the use of visual elements is contradictory. Brusilovsky and Miller (1999) suggest images may function as distractors, and Choi et al. (2003) argue that excluding visuals enhances validity. This contradiction highlights an ongoing lack of consensus on the optimal multimedia design in CBTs and its impact on the measured listening construct. Similarly, while CBTs allow for interactive and innovative items in theory (Parshall et al., 2009), the evidence base on their construct validation compared to traditional item types (like the multiple-choice items often favoured for practicality, Weir, 1988) remains limited.

2.2. Test anxiety and computer-based tests

Test anxiety, defined as the mixture of worry, fear, and discomfort arising from testing situations (Grupe & Nitschke, 2013; Putwain, 2008), is a critical factor that affects test-taker performance (Srisinthon, 2024). The transition to CBT raised immediate concerns regarding whether the computer interface itself would introduce mode-specific anxiety, particularly among test-takers with low computer familiarity.

Early research highlighted a significant contradiction in findings that appears largely linked to the novelty of computer administration at the time of the study. The early study by Ward et al. (1989), which assigned college students to computerized or traditional formats, found a significant difference with higher anxiety reported by the CBT group. This difference was attributed specifically to the novelty and unfamiliarity of the administration mode during that era.

In contrast, later studies, such as Chin et al. (1991), found no statistically significant variance in test anxiety between the PPT and CBT groups, even though the CBT group scored higher. The authors suggested that simple, well-structured CBT designs might mitigate anxiety concerns, even among students with limited prior exposure.

This shift in findings suggests a limitation in applying early research conclusions as computer familiarity has become almost universal, mode-specific anxiety caused by technology exposure is decreasing. More recent research supports this trend, with Khoshshima et al. (2019) finding no significant correlation between anxiety levels and the mode of administration among Iranian EFL test-takers.

Despite the general conclusion that computer familiarity eliminates the anxiety effect (Fazeli et al., 2013), there remains a gap in understanding the specific cognitive or affective components of anxiety that might still be triggered by CBT features. Some findings suggest a marginal, though not statistically significant, benefit to the computerized mode. Powers (2001) and Fritts and Marszalek (2010) both concluded that test anxiety was marginally lower among CBT examinees compared to their paper-based counterparts, especially after controlling for pre-existing trait anxiety.

The field currently lacks critical evaluation regarding why this marginal reduction occurs. Is it due to the structure of CBT with features such as automated time management and easier navigation or simply the normalization of digital interfaces? Future research must move beyond those comparisons to investigate the effect of specific interface elements like visible countdown timers and complex navigation interfaces on test anxiety (Zeidner, 1998; Ormrod & Davis, 2004).

2.3. Technical challenges and operational issues

The implementation of Computer-Based Testing (CBT) presents a distinct set of technical and logistical trade-offs that challenge assessment standardization and administration.

An advantage of CBT is the potential for cost mitigation over time, specifically by curtailing variable costs associated with traditional PPTs like printing, shipping, and storage (Coles et al., 2007; Drasgow, 2005). However, this often overlooks the significant limitation of the required investment needed for hardware implementation, system development, and software licensing (Arce-Ferrer & Martinez Guzman, 2009). The

financial burden presents a significant operational challenge that may prevent smaller institutions from adopting the mode.

Similarly, while CBT promises improved test security (Parshall et al., 2002), this gain is threatened by another increase in digital risks. CBT necessitates the deployment of highly sophisticated security software to guard against unauthorized duplication and distribution, making the integrity of the test reliant on the continuous upkeep and resilience of digital infrastructure (Maynes, 2009).

A major operational gap highlighted in the literature is the unpreparedness of institutional infrastructures. Several researchers express concern over the lack of readiness among educational institutions to transform assessment systems (Kingston, 2008). This lack of standardization threatens the comparability of test outcomes across different administrations. In cases of technical failure such as unstable network connections, students have reported frustration. This demonstrates that external technical variables can significantly affect the test outcome (Kingston, 2008).

CBTs offer clear advantages in gathering rich process data, such as time taken per item and mouse movements, which are unattainable with PPTs and invaluable for analyzing test-taker behavior (Parshall et al., 2002). Furthermore, the mode enables test designers to integrate multimedia objects to measure constructs more realistically (Al-Amri, 2009; Parshall et al., 2009).

However, the effective use of these features is constrained by technology. For example, the limitation on the amount of material that can be simultaneously displayed on a screen, compared to a sheet of paper, can restrict the presentation of questions and response areas (Bridgeman et al., 2001). Monitor size and resolution may also influence student preference and performance.

In listening tests, the necessary use of high-quality input material and sound editing equipment poses a technical complexity (Buck, 2001). This necessity helps explain why research focusing on listening test development often lags behind that of reading tests (Sawaki, 2001).

3. Method

The study used a mixed methods approach for data collection and analysis. This approach, which uses both qualitative and quantitative methods, strengthens the obtained data and addresses potential weaknesses that could result from relying on only one approach (McMillan, 2016). The impact of Computer-Based Tests (CBTs) on test-taker performance was studied quantitatively, while the impact on test-taker anxiety was examined both quantitatively via questionnaires and qualitatively through interviews.

3.1. Participants and settings

The subjects for the study were selected from a private university in Alexandria, Egypt. The students are enrolled in an English language course which covers the four language skills in addition to vocabulary and grammar. This university was selected because of the accessibility to computer labs and the familiarity of students with computer-based testing which is used at the initial stages of the student enrollment; this was a necessary prerequisite for the implementation of the study. At the time of the study, the researcher was the assistant-director of academic affairs and the head of the testing unit.

The sample consisted of 40 upper-intermediate English language university students who were randomly assigned to two groups: the experimental group ($n = 20$) and the control group ($n = 20$). All participants shared a highly similar educational and socioeconomic background, having graduated from the national general secondary education system before entering the university. This homogeneity ensured that baseline educational exposure was consistent across the sample. All students had a similar level of computer experience. They had successfully completed one mandatory general Information and Communication Technology (ICT) course provided by the university. They were familiar with computerized testing, since they have completed their university placement test in a computerized format. This prior experience was controlled to mitigate the impact of technology novelty on the findings. Furthermore, all study participants passed the previous English language level which guarantees baseline equivalence among them.

3.2. Research design

The study utilized the two-group posttest-only design. In this design, a random sample is divided into two groups: an experimental group that is exposed to the intervention (CBT administration) and a control group that is not (PPT administration) (Allen, 2017). The primary aim of this design was to investigate the impact of computerized reading and listening tests on the test performance and test anxiety levels of the upper-intermediate students. Using randomly assigned subjects who share a common educational and computer background enabled the investigation to reveal the extent to which comparability can be achieved across the two testing modes on the same measured constructs.

3.3. Measures and instruments

The researcher designed three listening comprehension tests based on Cambridge specifications which are followed in the English Language Centre. Each test comprises six short dialogues each with a context sentence, as well as a multiple-choice question or a sentence to complete and three alternatives.

To test reading, three reading comprehension tests were designed by the researcher according to Cambridge specifications which are followed in the language centre. Each test consists of five short texts with each text followed by a multiple-choice question. Students should use the visual information such as the layout to help identify the context and should also look for details in the text. Then, they answer the three-option multiple choice question after each text choosing only one answer. The task tests the students' general understanding and skimming skills of various kinds of short texts.

Using multiple tests allows for greater coverage of the content specifications. This minimizes the impact of a single component on the overall results, thereby enhancing the content validity and reliability of the assessment across the groups. Since the study aims to assess comparability across modes (CBT vs. PPT), it was essential to demonstrate that the findings were not dependent on one specific set of test items. By administering three equivalent tests to both groups, the researcher aimed to show that the measured mode effect remained consistent across different tests.

The researcher used an adapted version of the Test Anxiety Inventory (Powers, 2001) to measure the students' anxiety levels. TAI is a 20-item self-report scale that is used to measure test anxiety in college and high school students. It is a 4-point Likert

scale and students respond to each statement by one of the four options: (1) Almost Never, (2) Sometimes, (3) Often and (4) Almost Always. It measures individual differences in test anxiety as a situation-specific trait (Spielberger, 1980). It is regarded as a concise and uncomplicated tool that the majority of high school and college students can typically complete within 10 minutes or less (Taylor & Deane, 2002). It comprises three subscales: Test Anxiety-Worry (TAI-W), which consists of 8 items, Test Anxiety-Emotionality (TAI-E), which has eight items as well, and Test Anxiety-Total (TAI-T), which is made up of the sixteen items in the two previously mentioned subscales in addition to four other items.

Several studies indicated that TAI enjoys a high test-retest reliability of .80 in students at high-schools and higher education, and high internal consistency reliability in each of the two subscales (0.91 for the TAI-W and 0.91 for the TAI-E). The total scale (TAI-T) also has a high internal consistency reliability estimate of 0.96. The TAI is considered a highly reliable instrument and by far the most used test anxiety scale (Rana & Mahmood, 2010; Szafranski et al., 2012).

An interview was used to add a qualitative aspect of the study and to achieve data triangulation to increase the reliability and the validity of the study and the collected data. It was also used to acquire unique personalized information about how students saw the study tests in terms of the mode of administration (Cohen et al., 2002). It is also a way of complementing and overcoming the drawbacks of the used instruments (Brown, 2001). A semi-structured interview consisting of three open-ended questions was developed by the researcher by using the literature and his experience to use in interviews.

The interview involved three questions with questions 1 and 2 aiming to investigate the anxiety levels of the study participants before and after the tests while question 3 seeks to find the participants' preferable testing mode.

The interview questions were:

Q1. How did you feel before the tests?

Q2. How did you feel after the tests?

Q3. Do you think that you could have received better scores if the tests had been computer-based / paper-based? (for the CBT group, paper-based and vice versa)

The interview was first reviewed by two faculty members experienced in qualitative research and corrected according to the feedback received from them.

3.4. Procedure

After getting necessary approvals from the university, the researcher discussed the study aims with the two classes and obtained informed consent from each study participant. As shown in Table 1, the researcher conducted the three reading tests on the last weekday of weeks 3, 5 and 7 and the listening tests in weeks 4, 6 and 8 of spring 2021 – 2022 semester. The tests were collected and graded either manually or automatically depending on whether they were PPTs or CBTs. The scores for each group were kept on an excel sheet with all student data secured. In designing the computer-based tests, every attempt was made to ensure that the test-taking flexibility was as equivalent as possible to the conventional paper-and-pencil tests. Examinees were given 15 minutes to answer the test and could move back and forth within the tests by scrolling up and down respectively, items could be skipped as well as answered in any order, and responses could be reviewed and changed any number of times.

On the first day of week 9, the students were given the TAI questionnaire and were asked to fill it in based on their experience. Again, the responses were recorded on an Excel sheet to make it easier to analyse later. This timing effectively captures the residual test-related anxiety of the testing experience. This focuses on whether the test mode (CBT vs. PPT) left a lasting negative impression that followed the experience. Given the need to administer the questionnaire to all groups simultaneously while navigating institutional timetables, this schedule allowed for controlled data collection while maintaining the measurement window as short as possible to ensure the relevance of the self-reported scores to the preceding test event.

Table 1
Timeline of study procedures

Week	Procedure
3	Reading test 1
4	Listening test 1
5	Reading test 2
6	Listening test 2
7	Reading test 3
8	Listening test 3
9	TAI questionnaire
10	Test anxiety Interview

Finally, in week 10, the researcher met with eight members from the control group and eight members from the experimental group to conduct the test anxiety interview. The responses were digitally recorded, and the researcher made sure that they were secured. At the end of the study, all 40 participants were awarded certificates of appreciation for their willingness to participate.

4. Results

4.1. RQ1: What is the impact of computer-based language testing on upper-intermediate university students' performance in listening comprehension?

The first research question aimed at investigating the impact of computer-based tests on students' performance in listening comprehension tests and the first null hypothesis stated that there was no statistically significant difference between the mean scores of the listening tests between control and experimental groups. The researcher conducted the t-test for independent samples along with the effect size. No significant difference was found in the mean scores and the standard deviation of the test scores between the two groups according to the results shown in Table 2. Therefore, changing the test administration mode neither had a positive nor a negative impact on the students' listening test performance.

4.2. RQ2: What is the impact of computer-based language testing on upper-intermediate university students' performance in reading comprehension?

The second research question sought to investigate the impact of computer-based tests on students' performance in reading comprehension tests and the second null hypothesis stated that there was no statistically significant difference between the mean scores of the

reading tests between control and experimental groups. After conducting the *t*-test for independent samples, the results did not show a significant difference in the mean scores and standard deviations for the reading test scores between the two groups. This indicates that the change of testing mode did not have a significant impact on the experimental group's test performance, as shown in Table 3.

Table 2

Independent-sample *t*-test for the listening test scores in the PPT and the CBT groups

	Group	<i>t</i>	<i>df</i>	Sig. 2-tailed
Test 1	Control	-0.9	38.0	0.4
	Experimental			
Test 2	Control	-1.1	38.0	0.3
	Experimental			
Test 3	Control	-0.2	38.0	0.9
	Experimental			

Table 3

Independent-sample *t*-test for the reading test scores in the PPT and the CBT groups

	Group	<i>t</i>	<i>df</i>	Sig. 2-tailed
Test 1	Control	-0.6	38.0	0.6
	Experimental			
Test 2	Control	-0.9	38.0	0.4
	Experimental			
Test 3	Control	-0.6	38.0	0.5
	Experimental			

In conclusion, the results in the table suggest no statistically significant difference in the students' reading test performance between the control and the experimental groups. Therefore, the null hypothesis H0 (2): There is no statistically significant difference at ($p \leq 0.05$) between the mean scores of the reading test between control and experimental groups was accepted and the alternative hypothesis was rejected.

4.3. RQ3: What is the impact of computer-based testing on upper-intermediate university students' anxiety levels?

The third research question was directed to examining the effect on computer-based tests on students' test anxiety levels and the third null hypothesis stated that there was no statistically significant difference with regard to the test anxiety levels of the control and experimental groups due to changes in test administration modes. The students were given the test anxiety inventory questionnaire (TAI) and a semi-structured interview was conducted to investigate the differences in the anxiety levels between the two groups.

The results were presented in Table 4 and Fig. 1, showing significant differences between the two groups in text anxiety.

4.4. Test anxiety interviews

A semi-structured interview consisting of two open-ended questions and one yes/no question was developed by the researcher in light of the related literature. Due to limitations in availability and willingness to participate, the researcher resorted to

purposive sampling in selecting the participants for the interviews. Maximum variation strategy was implemented by choosing the top four and bottom four members from each group based on their mean scores in reading and listening tests (Palinkas et al., 2015). An interview was first conducted with eight members from the PPT group and then another interview was conducted with eight members from the CBT group. The interview involved three questions with questions 1 and 2 aiming to investigate the anxiety levels of the study participants before and after the tests while question 3 seeks to find the participants' preferable testing mode.

Table 4

Independent-sample *t*-test for the TAI scores in the PPT and the CBT groups

	Group	<i>t</i>	<i>df</i>	Sig. 2-tailed
TAI-W	Control	-6.210939	38.0	2.9248E-7
	Experimental			
TAI-E	Control	-6.609682	38.0	8.3364E-8
	Experimental			
TAI-T	Control	-8.993595	38.0	5.9751E-11
	Experimental			

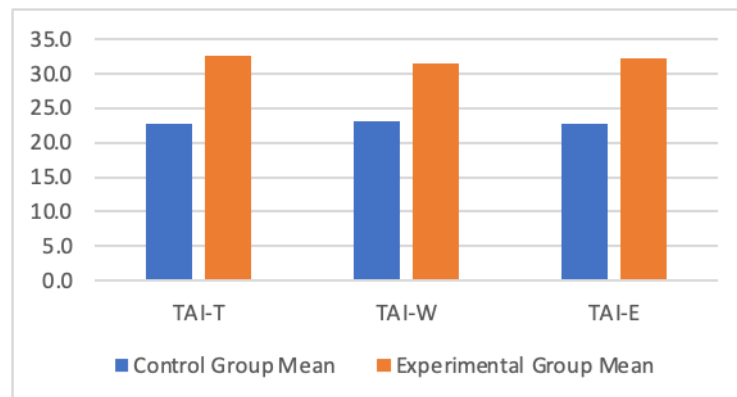


Fig.1. The mean scores of TAI-T, TAI-W & TAI-E

The qualitative analysis of the results was done using MAXQDA which gives specific codes to each response type. The open-ended responses to Questions 1 and 2 were analyzed using an inductive thematic analysis approach. The researcher read the transcripts thoroughly to identify recurring ideas, phrases, and concepts related to anxiety. Specific codes were assigned to segments of text within MAXQDA. These initial codes were then grouped into larger categories to understand the patterns of responses. The categorized responses were matched back to the corresponding participant's assigned group to identify impacts of the mode of administration.

To enhance the credibility of the qualitative findings, inter-rater reliability (IRR) was established for the coding process. A second, independent researcher was trained on the coding scheme and provided with a subset of the transcripts. Both the primary researcher and the independent coder applied the coding scheme to this subset. The agreement between the two raters was calculated using percent agreement (90%) indicating a high level of consistency in the interpretation and application of the codes. All discrepancies were resolved through discussion to finalize the coding framework. The

results showed that the CBT group had more anxiety levels than the PPT group before and after the tests. They also showed that almost one third of the CBT group members believed their scores could have been better if the tests were on paper, unlike the PPT group participants who all agreed that their marks could not have been better if the tests had been computerized. Therefore, the results of the interview show conformity with the results of the questionnaire.

Table 5

Statistical analysis of the responses to the interview questions

	Control group		Experimental group	
	Worried	Not Worried	Worried	Not Worried
Q1	25%	75%	50%	50%
Q2	12.5%	87.5%	25%	75%
Q3	0%	100%	37.5%	62.5%

The results (see Table 5) show that the experimental group interviewees have shown higher anxiety levels prior to the tests compared to their counterparts from the control group. This may be due to the fact that the majority of the tests they get are in paper-and-pencil mode, so familiarity plays an important role in test-taker anxiety when it comes to computer-based tests. It is also evident that the test anxiety levels declined in the control group after the tests while in the case of the experimental group, the results show less anxiety compared to the results relating to the anxiety levels before the tests. However, those results still show higher anxiety levels than those in the control group which supports the results obtained from the TAI questionnaire. The answers to the third question show that the interviewed members from the control group expressed higher preference of the PPT while 37.5% of the experimental group highlighted that would rather take the tests in paper-and-pencil than computer-based form.

5. Conclusion and recommendations

In light of the growing integration of technology in language assessment, this research examined the comparability of Computer-Based Testing (CBT) and Paper-Based Testing (PBT) in reading and listening tests, exploring potential impacts on test performance and test-taker anxiety levels. The findings from the score analysis indicated no noteworthy disparities in test scores between the two modes. However, the administration of a test anxiety questionnaire and subsequent interviews revealed that participants in the control group exhibited lower levels of anxiety compared to those in the experimental group. Further analysis of interview data unveiled a preference among more students for PBT, with CBT mode being associated with heightened test anxiety.

The results of this study suggest that computer-based testing (CBT) does not have a significant impact on test performance. However, CBT was found to significantly increase test anxiety. This finding is consistent with previous research, which has shown that computers can be a source of anxiety for some people. Those results conform to recent literature suggesting that changes in testing mode is not likely to impact the test performance (Yao, 2019; Yu & Iwashita, 2021; Sangsuwan & Rukthong, 2023). With regards to the impact of changing the testing mode on test-taker anxiety, the study results show similar results to similar studies (Shermis & Lombard, 1998; Colwell, 2013).

There are a number of possible explanations for the increase in test anxiety associated with CBT. First, computers can be unfamiliar and intimidating for some people. Second, CBT often requires test takers to interact with the computer in a variety of ways, such as clicking on choices or dragging and dropping text. This can be challenging for some people and can lead to feelings of frustration and anxiety. Third, because of its abstract nature, some students find it easier to work with paper and pencil rather than computers, which can add to the pressure and anxiety felt by test takers. Anxiety often stems from design issues, such as the constant presence of a countdown timer, difficulty with navigation and review, or the lack of immediate confirmation that answers are saved. Furthermore, environmental factors, like the fear of system crashes or network failure may increase worry. Additionally, providing dedicated technical support and orientation sessions.

The findings of this study have a number of implications for the use of CBT in reading and listening tests. First, it is important to be aware of the potential increase in test-taker anxiety when using CBTs due to the fact that it is not the usual testing mode that the students are exposed to. Second, it is important to take steps to reduce test anxiety by providing training on how to use the computer and how to interact with the test can increase comfort and build confidence in the system's reliability. Moreover, to reduce this anxiety in future trials, CBT interfaces must be modified to allow for more flexibility and offer enhanced visual confirmation of saved work.

The study's conclusions must be interpreted within specific limitations, primarily that the findings are specific to this homogeneous university sample and that the anxiety measure captured residual anxiety rather than acute state anxiety. Future research should address these gaps by investigating the impact of CBT on constructed-response items and utilizing real-time physiological measures to capture test anxiety instantly after the tests.

Overall, the results of this study suggest that CBT is a viable alternative to paper-and-pencil testing and a feasible solution to respond to the growing needs of students and educational institutions. However, it is important to be aware of the potential for increased test anxiety associated with CBT. By taking steps to reduce test anxiety, test managers can help ensure all students have a positive testing experience.

Author Statement

The author declares that there is no conflict of interest.

Acknowledgements

The author would thank Professor Maysoun Saad Saleh Omar and Dr. Maha Mohamed El-Komy for their valuable guidance and support. The author is also grateful to PUA's higher management for facilitating the implementation of this research on the university premises.

ORCID

Andrew Ghaly  <https://orcid.org/0009-0000-6679-5385>

References

- Al-Amri, S. (2008). Computer-based testing vs. paper-based testing: A comprehensive approach to examining the comparability of testing modes. *Essex Graduate Student Papers in Language & Linguistics*, 10, 22–44.
- Al-Amri, S. (2009). *Computer-based testing vs. paper-based testing: Establishing the comparability of reading tests through the revolution of a new comparability model in a Saudi EFL context*. Doctoral dissertation, The University of Essex.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
- Allen, M. (Ed.). (2017). *The SAGE encyclopedia of communication research methods*. SAGE Publications.
- Arce-Ferrer, A. J., & Martínez Guzmán, E. (2009). Studying the equivalence of computer-delivered and paper-based administrations of the raven standard progressive matrices test. *Educational and Psychological Measurement*, 69(5), 855–867. <https://doi.org/10.1177/0013164409332219>
- Boo, J. (1997). *Computerized versus paper-and-pencil assessment of educational development: Score comparability and examinee preferences*. The University of Iowa.
- Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2001). Effects of screen size, screen resolution, and display rate on computer-based test performance. *ETS Research Report Series*, 2001, i–23. <https://doi.org/10.1002/j.2333-8504.2001.tb01865.x>
- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge University Press.
- Brusilovsky, P., & Miller, P. (1999). Web-based testing for distance education. In *Proceedings of WebNet World Conference on the WWW and Internet* (pp. 149–154).
- Buck, G. (2001). *Assessing Listening* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511732959>
- Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28(3), 282–299. <https://doi.org/10.1080/08886504.1996.10782166>
- Chin, C. H. L., Donn, J. S., & Conry, R. F. (1991). Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. *Educational and Psychological Measurement*, 51(3), 735–745. <https://doi.org/10.1177/0013164491513025>
- Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295–320. <https://doi.org/10.1191/0265532203lt258oa>
- Cohen, L., Manion, L., & Morrison, K. (2002). *Research Methods in Education* (5th ed.). Routledge. <https://doi.org/10.4324/9780203224342>
- Coles, M. E., Cook, L. M., & Blake, T. R. (2007). Assessing obsessive compulsive symptoms and cognitions on the internet: Evidence for the comparability of paper and Internet administration. *Behaviour Research and Therapy*, 45(9), 2232–2240. <https://doi.org/10.1016/j.brat.2006.12.009>
- Colwell, N. M. (2013). Test anxiety, computer -adaptive testing, and the common core. *Journal of Education and Training Studies*, 1(2), 50–60. <https://doi.org/10.11114/jets.v1i2.101>
- Drasgow, F. (2005). The work ahead: A psychometric infrastructure for computerized adaptive tests. In Mills, C. N., Potenza, M. T., Fremer, J. J., & Ward, W. C. (Eds.), *Computer-Based Testing* (1st ed., pp. 1–35). Routledge.
- Epstein, J., & Klinkenberg, W. D. (2001). From Eliza to Internet: A brief history of computerized assessment. *Computers in Human Behavior*, 17(3), 295–314. [https://doi.org/10.1016/S0747-5632\(01\)00004-8](https://doi.org/10.1016/S0747-5632(01)00004-8)
- Fazeli, P. L., Ross, L. A., Vance, D. E., & Ball, K. (2013). The relationship between computer experience and computerized cognitive test performance among older

- adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(3), 337–346. <https://doi.org/10.1093/geronb/gbs071>
- Fritts, B. E., & Marszalek, J. M. (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychology of Education*, 13(3), 441–458. <https://doi.org/10.1007/s11218-010-9113-3>
- Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, 14(7), 488–501. <https://doi.org/10.1038/nrn3524>
- Hu, G., & Cao, R. (2025). A study on the digital transformation of higher education in Egypt. *Journal of Contemporary Educational Research*, 9(2), 101–105. <https://doi.org/10.26689/jcer.v9i2.9748>
- Johnson, M., & Green, S. (2006). On-line mathematics assessment: The impact of mode on performance and question answering strategies. *Journal of Technology, Learning, and Assessment*, 4(5).
- Khoshsim, H., Toroujeni, S. M. H., Thompson, N., & Ebrahimi, M. R. (2019). Computer-based (CBT) vs. paper-based (PBT) testing: Mode effect, relationship between computer familiarity, attitudes, aversion and mode preference with CBT test scores in an Asian private EFL context. *Teaching English with Technology*, 19(1), 86–101.
- Kingston, N. M. (2008). Comparability of computer- and paper-administered multiple-choice tests for K–12 populations: A synthesis. *Applied Measurement in Education*, 22(1), 22–37. <https://doi.org/10.1080/08957340802558326>
- Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6(1), 1–24. https://doi.org/10.1207/s15327574ijt0601_1
- MacCann, R. (2006). The equivalence of online and traditional testing for different subpopulations and item types. *British Journal of Educational Technology*, 37(1), 79–91. <https://doi.org/10.1111/j.1467-8535.2005.00524.x>
- Mayes, D. K., Sims, V. K., & Koonce, J. M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, 28(6), 367–378. [https://doi.org/10.1016/S0169-8141\(01\)00043-9](https://doi.org/10.1016/S0169-8141(01)00043-9)
- Maynes, D. (2009). *Caveon speaks out on IT exam security*. Retrieved from <https://caveon.com/resource/caveon-speaks-out-on-it-exam-security/>
- McKee, L. M., & Levinson, E. M. (1990). A review of the computerized version of the self-directed search. *The Career Development Quarterly*, 38(4), 325–333. <https://doi.org/10.1002/j.2161-0045.1990.tb00222.x>
- McMillan, J. H. (2016). *Fundamentals of educational research* (7th ed.). Pearson.
- Ministry of Education. (2014). *Strategic plan of pre-university education 2014 - 2030: Providing quality education for every child*. Ministry of Education, Cairo. Retrieved from <https://www.unesco.org/education/edurights/media/docs/c33b72f4c03c58424c5ff258c66aeace0eb58de4.pdf>
- Noyes, J. M., & Garland, K. J. (2003). VDT versus paper-based text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics*, 31(6), 411–423. [https://doi.org/10.1016/S0169-8141\(03\)00027-1](https://doi.org/10.1016/S0169-8141(03)00027-1)
- Ormrod, J. E., & Davis, K. M. (2004). *Human learning* (pp. 1–5). Merrill.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>

- Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. J. (2009). Innovative items for computerized testing. In W. J. Van Der Linden & C. A. W. Glas (Eds.), *Elements of Adaptive Testing* (pp. 215–230). Springer. https://doi.org/10.1007/978-0-387-85461-8_11
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). Considerations in computer-based testing. In C. G. Parshall, J. A. Spray, J. C. Kalohn, & T. Davey, *Practical Considerations in Computer-Based Testing* (pp. 1–12). Springer. https://doi.org/10.1007/978-1-4613-0083-0_1
- Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *The Journal of Technology, Learning and Assessment*, 2(6).
- Pomplun, M., & Custer, M. (2005). The score comparability of computerized and paper-and-pencil formats for K-3 reading tests. *Journal of Educational Computing Research*, 32(2), 153–166. <https://doi.org/10.2190/D2HU-PVAW-BR9Y-J1CL>
- Pomplun, M., Frey, S., & Becker, D. F. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62(2), 337–354. <https://doi.org/10.1177/0013164402062002009>
- Powers, D. E. (2001). Test anxiety and test performance: Comparing paper-based and computer-adaptive versions of the graduate record examinations (GRE) general test. *Journal of Educational Computing Research*, 24(3), 249–273. <https://doi.org/10.2190/680W-66CR-ORP7-CL1F>
- Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties*, 13(2), 141–155. <https://doi.org/10.1080/13632750802027713>
- Rana, R. A., & Mahmood, N. (2010). The relationship between test anxiety and academic achievement. *Bulletin of Education and Research*, 32(2), 63–74.
- Sandene, B., Horkay, N., Bennett, R., Allen, N., Braswell, J., Kaplan, B., & Oranje, A. (2005). *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project, research and development series*. U.S. Government Printing Office, U.S. Department of Education, National Center for Education Statistics. Retrieved from <https://files.eric.ed.gov/fulltext/ED485780.pdf>
- Sangsuwan, W., & Rukthong, A. (2023). Test-takers' performances on and perceptions of two different modes of online speaking tests. *LEARN Journal: Language Education and Acquisition Research Network*, 16(2), 168–183.
- Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*, 5(2), 38–59.
- Scheuermann, F., & Björnsson, J. (2009). *The transition to computer-based assessment: new approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities.
- Shermis, M. D., & Lombard, D. (1998). Effects of computer-based test administrations on test anxiety and performance. *Computers in Human Behavior*, 14(1), 111–123. [https://doi.org/10.1016/S0747-5632\(97\)00035-6](https://doi.org/10.1016/S0747-5632(97)00035-6)
- Spielberger, C. D. (1980). *Preliminary professional manual for the test anxiety inventory*. Consulting Psychologists Press.
- Srisinthon, P. (2024). Anxiety and engagement in the online classroom: A case study of Thai learners of Chinese as a foreign language. *Knowledge Management & E-Learning*, 16(4), 716–735. <https://doi.org/10.34105/j.kmel.2024.16.033>
- Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. *Anxiety, Stress, & Coping*, 25(6), 667–677. <https://doi.org/10.1080/10615806.2012.663490>

- Taylor, J., & Deane, F. P. (2002). Development of a short form of the Test Anxiety Inventory (TAI). *The Journal of General Psychology*, 129(2), 127–136. <https://doi.org/10.1080/00221300209603133>
- Ulfa, S., Surahman, E., Wedi, A., Fatawi, I., & Bringula, R. (2025). An adaptive assessment: Online summary with automated feedback as a self-assessment tool in MOOCs environments. *Knowledge Management & E-Learning*, 17(1), 88–113. <https://doi.org/10.34105/j.kmel.2025.17.004>
- Wannas, A. S., & AbdelMohsen, M. M. (2025). A systematic review of EFL online assessment in higher education: Effectiveness, attitudes, and challenges. *Knowledge Management & E-Learning*, 17(3), 435–453. <https://doi.org/10.34105/j.kmel.2025.17.020>
- Ward, T. J., Hooper, S. R., & Hannafin, K. M. (1989). The effect of computerized tests on the performance and attitudes of college students. *Journal of Educational Computing Research*, 5(3), 327–333. <https://doi.org/10.2190/4U1D-VQRM-J70D-JEQF>
- Weir, C. J. (1988). Communicative language testing with special reference to English as a foreign language. *Exeter Linguistic Studies*, 11, 1–241.
- Whitworth, B. (2001). *Equivalency of paper-and-pencil tests and computer-administered tests*. University of North Texas.
- Wise, S. L., & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development*, 23(1), 3–3.
- Yao, D. (2019). 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. <https://doi.org/10.5539/elt.v13n1p124>
- 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. <https://doi.org/10.1186/s40468-021-00147-0>
- Zeidner, M. (1998). *Test anxiety: The state of the art*. Plenum Press.