



The Impact of Experience and Academic Degree on EFL Test-Takers' Perception of Computer-Assisted Language Testing

Sedigheh Hajizadeh

Ph.D. candidate of English Language Department, Razi University, Kermashan, Iran.

sedigheh.hajizadeh2015@gmail.com

Vahid Reza Mirzaeian (Corresponding Author)

Assistant Professor of English Department, Faculty of Literature, Alzahra University, Tehran, Iran.

mirzaeian@alzahra.ac.ir

Mohsen Shirazizadeh

Faculty of Language Studies at Sohar University, Oman.

mohsenshirazizadeh@edu.om

ARTICLE INFO:

Received date:

2026.05.04

Accepted date:

2026.06.16

Print ISSN: 2251-7995

Online ISSN: 2676-6876

Keywords:

CALT, Experience, Online Testing, Evaluation Methodologies, Information Literacy.



Abstract

Assessment plays a vital role in educational systems. With the emergence of computer-based teaching, global attention has shifted from traditional to online teaching, but not necessarily to technology-based testing. As a new form, testing through technology requires careful consideration, as decisions based on test results may significantly impact examinees' futures. An exploratory mixed-methods design was used to examine examinees' perceptions of computer-assisted language testing (CALT) and its relationship to test-takers' academic degrees and experience. To this end, 162 Iranian students studying translation, teaching, linguistics, and English literature were recruited to complete an existing questionnaire, and 11 participants were selected from the 165 to participate in interviews. The data were gathered at Payam Nour University in Iran through an adapted questionnaire. A two-way ANOVA was employed to examine the effects of experience with technology-based testing and academic degree on testing results. Regarding the qualitative part, we used observation and semi-structured interviews to collect in-depth data, drawing on interpretation and reflection to meaningfully code and categorize the data. Finally, three themes emerged from the data through in vivo and open coding, using a bottom-up coding scheme. The results revealed that prior exposure to this technology significantly affected examinees' perceptions, regardless of their academic degree. The findings of this study can be beneficial to policymakers, administrators, and test developers in making informed decisions about whether and how to utilize this technology.

Citation: Hajizadeh, S.; Mirzaeian, V, R. & Shirazizadeh, M. (2026). The Impact of Experience and Academic Degree on EFL Test-Takers' Perception of Computer-Assisted Language Testing. *Journal of English Language Teaching and Learning*, 18 (37), 229-252. DOI: 10.22034/elt.2026.72214.2866

Introduction

In recent years, computers have become an integral part of language learning and testing worldwide, offering many opportunities while posing enormous demands (Chapelle & Douglas, 2006; Chapelle & Sauro, 2017; Chapelle & Voss, 2016). As the number of examinees has grown computer-assisted language testing (CALT) has been widely adopted internationally in academic contexts. With the advent of computer-based testing, many educational institutions have attempted to transition from paper-and-pencil tests (PPT) to computer-based tests (CBT) (Olufemi & Oluwatayo, 2014). Today's concern is no longer whether to employ computers for learning and testing; hence, many studies have focused on computer-related factors, their contributions, and their impacts on test performance to develop and implement them appropriately (Chan et al., 2018; Chapelle, 2003; Chapelle & Douglas, 2006; Chapelle & Sauro, 2017; Chapelle & Voss, 2016; Jin & Yan, 2017; Suvorov & Hegelheimer, 2014; Wise, 2019).

The majority of studies have concentrated on whether PPT and CBT were equivalent in terms of attitude (Tella & Bashorun, 2012), scores (Voss, 2018), computer experience or familiarity (Khoshshima et al., 2019), gender (Taylor et al., 1999), and Computer Anxiety (CA) (Balogun & Olanrewaju, 2016). Since technology is being employed and will continue to be employed for language testing (Douglas, 2012), administrators should be cautious about the effects of the test method (Chapelle, 2003) to maximize the advantages of computers and mitigate their unwanted consequences. As such, some testing pioneers reported the relationship between the two modes and their potential impact on test performance (Bachman & Palmer, 1982).

Given that the “investigation of validity is both a matter of judgment and empirical research, it involves gathering empirical evidence and appraising the values and social consequences justifying specific interpretations or uses of test scores” (Bachman, 1990, 237). Decisions made based on test results can significantly impact examinees' performance due to any unsystematic errors caused by the test mode; hence, they require careful consideration (Bachman & Palmer, 2010). Therefore, it is crucial to investigate the challenges EFL test takers face when taking tests on computers and their perspectives on CALT to avoid unsystematic errors or construct-irrelevant variance.

Some testing experts have investigated the threats and promise associated with computerized tests by considering test-takers' perspectives on CALT (Chapelle, 2003). These perspectives can be rooted in the act of taking tests purely out of habit, namely, PPT, oral assessments, and interviews, or they can stem from a lack of computer access and literacy. Although the employment of CALT is evolving rapidly, and assessment tools and dynamic strategies are increasing (Bahari, 2020), to the best of our knowledge, the test-takers' perspectives based on their exam experience and academic degree towards CALT have not been adequately investigated. In other words, the examinees' perspectives and hindrances regarding CALT must be explored in terms of their actual meaning. Notwithstanding, Brown (2019) emphasized the importance of the experience of the manipulating system “to ensure equity of opportunity to learn how to interact with the system and to ensure the quality of

experience with the system, so that construct-irrelevant factors do not contaminate resulting scores” (p. 17).

Any construct-irrelevant variance and construct underrepresentation (Taylor et al., 1999) can cause test score inflation or deflation in favor of one group and against another, thereby threatening validity (Wise, 2019). The threats include “costs of a CBT testing program such as computer infrastructure, technical support, standardized testing environment, computer maintenance, and test takers’ psychometric issues, i.e., computer familiarity” (Wise, 2019, 22). Likewise, test takers’ computer familiarity or experience, as well as their perspectives toward test delivery (e.g., CALT), may confound score interpretation, and some unsystematic, irrelevant constructs may contaminate the measurement of testees’ language proficiency (Jin & Yan, 2017).

Since EFL perspectives are shaped by experience and influenced by external factors (Tella & Bashorun, 2012), test takers’ perspectives on the test mode are crucial for identifying how they can be improved and how the challenges can be removed to enhance test delivery (Tella & Bashorun, 2012). However, to the best of our knowledge, there is a paucity of research regarding test-takers’ perceptions of computerized exams in the Iranian context. As such, the study investigated whether exam experience or academic degree affected perceptions towards CALT. Likewise, it examined whether inexperienced and experienced graduate students felt more positively toward CALT than their inexperienced or experienced undergraduate counterparts. The study aimed to unfold the interaction between CALT experience and academic degrees. As such, we examined whether test takers’ perspectives can be influenced by their experiences taking the test on computers and by their academic backgrounds. In addition, the research examined whether a higher degree can influence their perspectives. Hence, qualitatively identifying Iranian students’ perceptions may reveal the voices of marginalized groups. The following research questions were formulated to guide the research:

1. What are the perspectives and dispositions of Iranian EFL instructors and test takers towards computerized tests in academic contexts?
2. Does the experience of taking exams on computers influence English students’ perspectives towards CALT?
3. Do the academic degrees (MA & BA) affect English students’ perspectives towards CALT at PNU?
4. Are there any differences in the perspectives of inexperienced graduates and undergraduates from those of experienced graduates and undergraduates?

Literature Review

Computerized language tests have evolved since the late 1980s, and test developers have expressed concerns regarding the suitability of test methods for particular contexts and examinees (Suvorov & Hegelheimer, 2014). Early CALT was delivered linearly, similar to its counterpart, PPT, which included multiple-choice, multiple-select, drag-and-drop, and short-answer responses, leading to a comparison of the CALT and PPT versions of the same test (Winke & Isbell, 2018). The review of empirical studies in this area revealed that CALT has

concentrated on comparisons between PPT and CALT (Bugbee, 1996; Dewi, 2025; Tella & Bashorun, 2012; Yeom & Jun, 2020).

It is argued that a lack of familiarity with the testing mode can significantly affect testees' performance. Some researchers have attempted to verify the equivalence of the two test-presentation modes by comparing examinees' performance on paper and CBT (Akdemir & Oguz, 2008; Khoshshima et al., 2019). However, mixed findings have been reported by comparing these two test modes. Although Hosseini and Toroujeni (2017) reported better performance on PPT in two reading comprehension tests, Khoshshima and Toroujeni (2017) found no significant difference between the two test modes with respect to gender and test preference. Therefore, the argument should be applied to various aspects of computerized tests. In another study, the test performance of EFL learners was equivalent across CBT and PPT modes (Khoshshima et al., 2019). Similarly, comparing computer-based and paper-and-pencil tests to determine whether unfamiliarity with the testing process on the exam affects students' performance, Burgmanis et al. (2023) found that results were the same across both versions.

The illustration of these two test delivery modes has revealed information about the construct being measured, although irrelevant constructs may also affect test performance. Notwithstanding, many individual measures with variable psychometric properties have been developed to assess these constructs (Schulenberg & Melton, 2008). Since no test method can ultimately measure a trait without introducing random or systematic errors, any changes may result in detrimental repercussions or rewarding outcomes that favor or disadvantage a particular group, and any source of construct-irrelevant variance (Taylor et al., 1999) threatens validity (Wise, 2019). Nevertheless, the most recent evaluations have focused on dynamic assessment and the concept of mediation, rooted in the Vygotskian perspective (Ebadi & Bashir, 2020). In essence, paradigm shifts in language testing methods have led to the integration of object regulation into CBT; test users employ artifacts as mediators to facilitate test evaluation (Lantolf et al., 2015). Adapting the concept of mediation beyond its original notion of psychological tools for computerized testing (Daniels, 2015), test developers have sought to facilitate the testing process through technology. Hence, they have been concerned with unsystematic measurement error and have sought to reduce it by investigating it in different ways.

One way to reduce unsystematic measurement errors is to identify them from the stakeholders' feelings toward CBT. For instance, Alrob et al. (2019) explored teachers' and students' perspectives on two test formats. They found that CBT results were significantly better than those of PPT. In another study, Jimoh et al. (2012) compared undergraduate students' perceptions of CBT, revealing that students complained that CBT could not evaluate their skills as effectively as the PPT mode did in their previous exams, while the test mode's immediate grading and speed were satisfactory. Although test-taking is regarded as an emotion-laden process that can negatively affect test-takers' emotions (Daniels & Gierl, 2017), familiarity with computers can reduce this burden. Investigating differences in scores between examinees on PPT and CBT, Chan et al. (2018) examined variations in test performance and the effect of computer familiarity on test performance. They found that a lack of familiarity with computers might lead to lower CBT performance among those without the required

knowledge. In a comprehensive review of 276 articles on CA from the 1990s and 2000s, Powell (2013) found that in some studies, the more experience the examinees gained, the less CA influenced their performance.

Given the importance of test familiarity, technology-related self-efficacy, and literacy in language testing, some tests have been developed to measure various aspects of construct-irrelevant digital demands. In this regard, Reviewing 105 studies on digital literacy, familiarity, and their relation to construct-irrelevant digital demands, Zhang and Razak (2026) found that unfamiliarity with the system can enhance anxiety during online assessment can reduce the performance. Some studies similarly emphasized the profound impact of computer familiarity on test performance, leading to better results accordingly. Concerning the effect of moderator variables, namely gender, computer literacy, and CA, Lee and Huang (2014) realized that “computer literacy contributed to positive computer self-efficacy and lower CA for males than females, and CA led to more negative computer self-efficacy and higher CA for females than for males”. In essence, a relationship existed between gender and CA. Most studies found that testees performed similarly across the two test modes; most examinees had difficulty with CBT and preferred taking the test on paper. In a recent study, Yeom and Jun (2020) compared reading tests administered on paper and on a computer, and found no significant difference in test-takers’ performance between the two modes. Namely, computerized tests can enhance both digital readiness and learning (Dewi, 2025). Although these studies showed that computer literacy had a negligible effect on the tests, they emphasized the need for computer tutorials to help test-takers gain experience with computers (Taylor et al., 1999).

Since computerized tests are not without their challenges (Brown, 2019), most stakeholders influenced by them exhibit concerns that test delivery may affect test performance (Chapelle, 2001). The test mode influences examinees’ performance, whether they obtain high or low scores; the difference between the two test delivery modes should be used as evidence of test validity (Chapelle, 2003). Therefore, in light of gathering evidence for construct validity (Bachman, 1990), many studies have explored the effect of test mode on test-takers’ performance to identify any construct-irrelevant variance (Messick, 1989, as cited in Taylor et al., 1999), contributing to systematic error variance in scores and thereby threatening validity (Wise, 2019). By doing so, researchers have attempted to identify any unexpected consequences arising from switching test modes, score deflation, and score inflation, which can lead to measurement error or unreliability (Bachman, 1990).

Regarding the status quo of computer-based language testing in the Iranian context, research on CALT has primarily focused on identifying affordances and drawbacks, or on comparing PPT and CBT in terms of computer familiarity and anxiety. Hosseini et al. (2014) explored EFL students’ attitudes toward CBT in Iran and found that familiarity with and attitudes toward computers had negligible effects on students’ performance. Moreover, they prioritized PPT over CBT. Similarly, Khoshsima et al. (2019) concluded that there was no significant effect of prior testing mode preference and gender on performance through CBT. Another study examined the effects of CBT on testees’ perspectives, and administrators at PNU demonstrated that, despite the limitations and advantages of computerized testing, most examinees and administrators hold positive views of CBT. Despite the affordances, the researchers argued that

CALT's limitations render it inefficient (Ockey, 2009), and these limitations need to be addressed.

Brown (2019) identified many challenges in developing a computerized testing system in any society. However, he argued that a comprehensive history of computerized testing is unavailable. He informed the Ministry of Education of plans to enhance the end-user experience and technology infrastructure to achieve the desired outcome. His study revealed that the quality of examinees' experience with the system is a crucial factor affecting scores. As examinees' perceptions are shaped by experience (Tella & Bashorun, 2012), the test mode is essential for universities to identify where improvements and obstacles lie in enhancing test delivery (Tella & Bashorun, 2012). However, little information has been revealed from test takers' viewpoints and experiences regarding EFL learners' perspectives on CALT in the Iranian context.

Method

Design of the Study

The current study employed an exploratory sequential mixed-methods design, beginning with the collection of qualitative data. The exploratory design was applied because the test takers' challenges and perspectives on their experiences needed to be qualitatively explored before being quantitatively measured. Namely, the qualitative data were collected and analyzed first, followed by the quantitative data collected and analyzed through a questionnaire.

To ensure that the degree and exam experience affected test takers' perspectives towards CALT, two independent variables were examined to determine their effects on examinees' performance. Moreover, in this investigation, the questionnaire has been adapted to incorporate both computer familiarity and computer perspectives in the items, as well as to address overall experiences. As such, a two-way ANOVA was employed in the current study because the data were normally distributed and lacked outliers.

Setting and Participants

The test delivery method employed at PNU is an offline exam system, where questions are stored on the server of a local area network (LAN). The examinees access the questions through the exam portal on computers by entering their passwords. Unlike adaptive tests, the test format delivered on computers at PNU is linear, allowing testees to review their responses, which is similar to the PPT version, i.e., adoption of PPT duplicates for CBT purposes (Bahari, 2020). The purpose of CALT employed at PNU is curriculum-related; summative exams measure students' achievement at the end of a unit and require learners to meet a criterion. Psychometric tests are employed at PNU, which is interested in objective tests, including multiple-choice and system-referenced tests in which only knowledge of the language system or linguistic elements is selected accordingly. The university's websites and posted posters provide instructions for examinees on how to log in to the exam portal.

The participants in the quantitative phase were students studying translation, teaching, linguistics, and English literature, aged 23 to 50 (see Table 1). The selection of participants was based on their experience with computerized tests and their academic degree. We

categorized groups as inexperienced (those who had taken computerized tests fewer than 5 times) and experienced (those who had taken them more than 5 times). The respondents were selected for their diversity, encompassing different age groups, genders, and fields of study, to provide relevant information from students with diverse English language backgrounds. Overall, 162 participants completed the questionnaire.

Table 1. *Demographic Characteristics of the EFL Prospective Teachers*

Characteristic	Subgroup	Percentage	N
Discipline	TEFL	60%	67
	Literature	25%	28
	Translation	15%	17
Gender	Female	70%	78
	Male	30%	34
Age	24–30 years	45%	50
	31–40 years	40%	45
	41–51 years	15%	17

In the qualitative study, the non-probability sampling technique of purposive sampling was employed for the interviews, which included participants of various ages and genders, to provide rich sources of information and elicit the testees' challenges and perspectives that may enhance or influence the test results. Overall, 11 examinees were interviewed from both experienced and inexperienced groups.

Instrument

The instruments were used in both qualitative and quantitative phases, including observations, interviews, and questionnaires, as explained below. Given the need for in-depth data, methodological triangulation was employed to enhance the credibility of the findings (Merriam & Tisdell, 2015).

Observation

This study used observation, a suitable data collection technique for monitoring behaviors that cannot be inferred from words alone. One of the researchers collected data over 12 days, including about 42 hours of exam observations. Although observing participants' behaviors and reactions while taking the test on computers can be a useful data-collection method, the researcher must be cautious, as the testing process is sensitive to any distractions. The purpose of the observation was to provide context and to derive specific data and behaviors that can serve as a reference for subsequent interviews.

Interview

In the qualitative phase, semi-structured interviews were conducted to collect examinees' experiences and perceptions of CALT. The second collection method employed in this study was a semi-structured interview, in which eleven test takers answered the questions individually. According to [Heigham and Croker \(2009\)](#), semi-structured interviews can take unexpected directions; namely, the interview is open to new questions. Accordingly, no fixed pattern was followed in this interview to elicit more information from the participants that could not be detected through observation. The interview is a perfect account of understanding phenomena from the participants' viewpoint, experiencing them in real situations.

Questionnaire

The data collection technique involved administering a predetermined set of closed-response questions and collecting numerical data. Among the various questionnaires, the most homogeneous ones relevant to the current research were selected and adapted or modified. In fact, the questionnaire was adapted from [Alrob et al. \(2019\)](#) and [Jimoh et al. \(2012\)](#), with minor changes to align with the research context. Reliability was assessed using Cronbach's Alpha, the most appropriate method. After checking reliability on a 5-point scale (1-5), the researchers reduced the 40 items to 37. The questionnaire was piloted with eight English students to identify potential ambiguities, ensure the questions were suitable for Iranian EFL students, and assess its validity.

Procedure

First, one of the researchers observed the examinees as they took their exams to identify any challenges with CALT. The researcher remained a complete observer, as it is impossible to play any role in the exam process. Since it was essential to ensure sufficient care to avoid diverting participants' attention, the researcher's presence did not distract the examinees from testing. Moreover, the collected data, along with the observer's comments and reflections, were recorded on the spot in field notes, with the aim that the observer would remember all the details later while analyzing and interpreting the data.

Drawing on observations, a list of predetermined interview questions was designed and piloted with one instructor and one student under the supervision of an applied linguistics professor. The instructors who had experience taking exams on computers were interviewed via face-to-face or online semi-structured interviews, in which preselected questions were asked first. Data collected through interviews were presented within the transcription of the words in which direct quotations were asked from participants about their experiences and feelings. It must be noted that this is a second-hand account of our investigation, based on the participants' interpretations and thoughts.

Following the observations and interviews, the adapted questionnaire was piloted with a group of EFL students to customize the questions for the target participants and prevent wasted time and energy. Pilot testing led to revisions and fine-tuning of the protocol, thereby avoiding the loss of valuable, potentially useful, and often irreplaceable data. The scale's internal consistency was examined to check the interrelatedness of items ([Pallant, 2011](#)). As mentioned by [Hinton et al. \(2004\)](#), "it is necessary for the questionnaire to measure the topic under study

at different times and across different populations” (p. 356). All items, including questions about computer anxiety, literacy, opportunities, and the limitations of CALT, were worded both positively and negatively, then reverse-scored and reverse-coded before checking for reliability (Larson-Hall, 2015; Pallant, 2011). Ultimately, the impact of two independent variables was checked. Then, the interaction effect of two factors was analyzed. The findings reported in the results and discussion section were then analyzed based on the main interaction effect and p-value.

Data Analysis

In the quantitative phase, the variables were evaluated using numerical data obtained from the questionnaire scores. Although the scores obtained from the questionnaires were not interval—the Likert scale is considered nonparametric in some studies—the parametric two-way ANOVA was employed to examine the main effects of the test-taker’s experience with the tests through CALT and academic degrees. The statistical test was chosen because the data were normally distributed and lacked outliers; thus, the parametric test assumptions were met. Moreover, the probable interaction effects between these two independent variables were examined.

To assess the effect of test-taking experience on computer reliability, the researchers measured internal consistency (Cronbach’s alpha). The results of the internal reliability coefficients are shown in Table 2.

Table 2. *Reliability Statistics*

Cronbach’s Alpha	Cronbach’s Alpha Based on Standardized Items	N of Items
.877	.878	37

The questionnaire’s reliability was confirmed after omitting three questions that did not meet the required criterion. The value was 0.877, as illustrated in the table, suggesting strong internal consistency or reliability for the scale.

This interpretive research aimed to understand the phenomenon from the multiple perspectives of different participants regarding CALT. Hence, we employed an inductive analysis procedure to interpret the phenomena by reducing and reconstructing the data into codes and categories. The researchers drew on interpretation and reflection of meaning (Richards, 2014) by reading and re-reading the transcripts several times. The themes emerged from the data through participants’ and researchers’ in vivo and open coding, using a bottom-up coding scheme. After defining codes, they became more similar; these reduced-meaning units were classified under their related categories. The categories were conceptual elements that “covered” many individual codes, which were abstractions derived from the data, not the data themselves (Heigham & Croker, 2009, 181). Finally, a few condensed meaning units with a related thematic core emerged as the final themes.

Results

Qualitative Phase

Computer Anxiety Experienced through CALT

Regarding anxiety arising from taking tests on low-grade computers, the primary deficiencies stemmed from hardware and software infrastructure, including poorly maintained systems. Unsystematic errors or unexpected repercussions in the process of test-taking were a source of distraction, and they led to anxiety. Using new systems with high-quality accessories, such as cutting-edge computers, reduces distractions, whereas low-grade equipment leads to examinees' dissatisfaction. The examinees complained about the quality of the computers and the problems they experienced while taking the tests. The low-grade computers led EFL students to harbor negative feelings about the safety of computerized exams. As some of the equipment and systems were not satisfactory, emotional stress, also known as computer anxiety, threatened test performance. For instance, one of the interviewees stated some flaws of the equipment as follows:

I believe facilities must be enhanced, as some keyboards and mice are rusty and make low-quality sound screens. I had to press twice with force to select the answers. We should not be worried about the computers; instead, we should focus on the test itself. We should focus on the exams, not on computer flaws. Taking the test by computer makes me worried, (Keyvan, an undergraduate student).

The absence of an auto-saving program was attributed to one of the software failures. When a system stops running and lacks an auto-save program, all data is lost. However, they can approach the test differently depending on how much time remains. In some situations, the program allows test takers to log in again, but this is not always possible. Otherwise, they take the exam on paper. Even though they manage to log in, they must repeatedly answer unsaved information under high stress.

I think most test takers' nightmare is the misfortune wherein all the answers disappear suddenly owing to whatever reasons. I want to finish the test as fast as possible without any interruption, and hate taking the test twice in one exam, (Samaneh, a graduate student).

When the computers stopped running for various reasons, such as a power outage caused by test takers unintentionally cutting off power to outlets or kicking the cases, all responses were lost due to the lack of an auto-save program. The program was flexible, allowing test takers to log in again as long as time remained. Otherwise, they went through taking the exam through PPT. However, all responses were lost. Notwithstanding, examinees managed to log in; they had to answer all questions again under high stress.

We do not forget the moment that the power went out unexpectedly. After waiting about two hours, we ended up taking the test through hard copies. We were tired, so we blew the test.

Some EFL students gathered in the cramped campus, waiting to enter the computer lab mentioned. Required technical support, proficient staff, the number of technical personnel, and their expertise to fix the problems must be sufficient so that the test takers feel secure during the test. Hence, they can thoroughly concentrate on the exam itself. Although most technical staff were competent and helpful, there was not enough technical support to assist two or three test takers simultaneously. Shima complained about her experience performing a test a couple of days ago as follows:

I was about to finish the test; suddenly, I unintentionally pulled the wire on my leg. I had to wait for help for a while since one of the personnel was busy with another system, and it took about 4 minutes for the technical staff to solve my problem.

Experience of Computerized Exam Creating Positive Perception

Although a new generation is born and raised in the technology era in many provinces in Iran, the test takers' computer literacy should not be taken for granted. Access to computers is limited in some remote areas, and some examinees "suffer from a lack of computer literacy" (Chapelle & Sauro, 2017). Since less experienced examinees complained that they received lower scores on the computerized tests, they believed the university should have provided instructions before the exams.

Administrators should facilitate a situation where enough experience is gained by doing and learning. I wouldn't say I liked taking the tests on a computer first. However, I keep my fingers crossed to take the exams by computer (Sara, a graduate student).

The information gleaned from the interviews and observations revealed that test takers with sufficient computer knowledge could confidently log in to the test portal and complete their exams without difficulty. More importantly, when any computer errors, like an uninvited guest, popped up during the exam, they did not panic and could fix them; otherwise, they asked for help without making a scene. Some even checked the wires and screen connections to ensure no cutoff would happen during the test.

I am not worried about problems caused by computers and only think about the exams; we have to be prepared for the materials, so I do not think about problems that may happen during the exams, (Sadegh, an undergraduate).

Some test takers thought the more experience and knowledge they acquired, the less anxiety they felt on their exams. Consequently, the more they took CALT, the more they enjoyed taking these tests. Ahmad, a graduate student, stated:

It was not easy for me to take my exam on a computer for the first time. While taking three tests on computers, I preferred them over PPT.

Observing the test takers, technical personnel, and proctors during the exam revealed some previously undiscovered points, such as the importance of technology literacy not only for technical personnel but also for examinees. The computer-literate examinees had greater self-confidence and a stronger predisposition toward CALT. Surprisingly, even most middle-aged

adults with average computer literacy were more comfortable using computers. As a result, they had enough self-confidence to perform the tests through the CALT. Hence, they were not anxious about performing their tests, given their prior experience and computer literacy.

Using technology at universities must be a routine, as university students are literate enough to use the computers for assessment. Technology literacy must be a prerequisite for students to have enough experience to manipulate the computers easily and to save time during the exams, (Mr. Safavi, an instructor).

Challenges vs. Opportunities of CALT

No new method is immune to criticism or resistance; Kessler and Hubbard (2017) assumed that CALT has been prone to criticism of new methods. Despite CALT having paved the way for a new mode of exam delivery, people resist new ideas and methods. Although technology has offered immense opportunities for learning and testing, it takes time for EFL learners to become accustomed to these new technologies; moreover, since CALT is relatively young in the Iranian context, most users are beginning to experience its advantages and disadvantages. The quote below was extracted from interviews about the novelty of CALT in the Iranian context, followed by some recommendations.

CALT is a culture where people resist new ideas and methods. It takes time for people to get used to these new technologies; however, we must move forward despite a lack of CALT acceptability by the authorities, computer literacy, sufficient equipment, and internet speed, (Dr. Tabatabaee).

Most of them claimed that their old habits of taking the exam through PPTs—such as underlining essential parts, making notes in the margins, or ruling out irrelevant sections—were among the other challenges.

I'm not used to taking exams by computer, as I like to hold the exam sheet and answer the questions with my pencil, and I do not feel comfortable reading on screens. What if I could underline, cross out, and circle the words that seem vague, and I have many doubts about them.

The test format led to under-representation of the construct due to the tests' negative washback. Multiple-choice tests, delivered every three or four tests, led examinees to skim some test series from former exams. Most students at PNU rely on past exams for studying and do not study the book at all, as the questions are drawn from a pool of exams that include repetitive, little-variant questions.

I usually study two nights before the exams to pass the test, even though I do not get a good mark, but a pass mark. I can buy the previous tests on the net or download them before the exams, (Ali, a graduate student).

Most PNU instructors stated that using selective task types cannot capture the multiple facets of the intended construct; therefore, the test methods cannot thoroughly measure it.

Although testing at PNU by computers was uniform through multiple-choice and highly reliable, the absence of innovation in the test method led to

assessing only discrete points and linguistic elements, but not a comprehensive assessment.

The first and foremost tremendous opportunity afforded by CALT is environmental conservation, driven by reduced paper consumption. CALL and CALT are environmentally friendly since no trees need to be cut down for single-use purposes. Consuming natural materials used just once for a test consumes natural resources, such as trees, and wastes money and energy in the preparation, transportation, distribution, copying, and collection of the test sheets. The technical staff and proctors emphasized the benefits of computerized tests.

The cost of applying PPTs is more expensive than CBT, owing to the distribution and returning of exam papers rather than electronically reading them, (two staff members).

Time-saving is one of the privileges of CALL and CALT, particularly in the learning and testing domains, where the accessibility of amenities and the economy of materials are made possible by technology. Most testees at PNU believe that CBT is faster than PPT and requires less energy due to the availability of computer facilities.

We panicked when it was announced that we had to take our exams by computer; we did not know what it looked like. After the second exam, we found out it was not a rocket job, and we finished the exam faster than PPT, (Ali, a graduate student).

Although the immediate test score had a positive influence on examinees' emotions and aspirations, it had adverse effects on anxiety and pride as well (Daniels & Gierl, 2017). Most examinees believed automatic and immediate scoring was soothing and enjoyable since it reduced after-exam anxiety. However, it also had some impact if the test results were not satisfactory or awkward, as the following example mentioned by an undergraduate student:

I failed today's test, so I had to retake this lesson and could not focus on tomorrow's exam.

By automatically storing data without the burden of carrying or sorting the test paper, PNU administrators applied the technology to save money, time, and energy.

Thank God! We do not have to gather paper, put them into order, and pile them for scoring, (One of the proctors).

Quantitative Phase

The results of the parametric two-way ANOVA, as applied to the second research question, were interpreted as follows. The mean score of experienced test-takers, both bachelor's and master's degree holders, was higher than that of inexperienced examinees. The inexperienced BA and MA groups had a mean of 134.0 (SD = 16.466), and the experienced BA and MA groups had a mean of 138.89 (SD = 15.73). The difference between the inexperienced and experienced groups indicated that the experienced groups had a more positive perspective towards CALT. Although the descriptive statistics revealed a difference, they did not indicate whether the difference was statistically significant. Thus, to fulfill this purpose, inferential

Discussion

This study aimed to investigate the CALT experience and academic degrees in relation to examinees' perspectives on computer-based tests. Using an exploratory mixed-methods design, we found that experience plays a paramount role in shaping testees' positive perspectives on CALT, regardless of academic degree. Namely, the first research question, related to the qualitative phase, revealed that more experienced examinees hold a positive perception of CALT, finding it promising and convenient. Regarding CALT experience, the second question showed that experience with computerized tests significantly affected test-takers' perceptions; i.e., the more experienced they were, the more positive their perceptions of CALT were. Regarding the third research question, the effects of a higher academic degree on examinees' perspectives towards CALT were insignificant; a higher degree did not affect the perceptions of experienced and inexperienced test-takers. The last research question showed no significant interaction effect between experience and degree.

In line with Jimoh et al. (2012), the key role of prior computer familiarity and experience, as well as the need for CBT training, was emphasized. Hence, the researchers suggested that CBT should be used in a few courses to familiarise testees with CBT and to customize the test to the subject matter and its requirements, which aligns with our research. However, the current study uncovered the positive perception of CALT. In contrast, the study by Jimoh et al. (2012) revealed that CBT improved examinees' performance, whereas a lack of CBT knowledge was associated with CA. On the other hand, regarding computer familiarity, experience, and testing mode preference, Burgmanis et al. (2023) found that these factors had a trivial impact on students' performance on CBT. They reported that CBT results were higher than those on the paper-and-pencil test, regardless of the test delivery mode.

Regarding computer experience and test performance, a lack of computer experience was associated with lower scores (Lee, 1986) and less confidence in CBT. Although our research design differed from Lee's, both studies revealed that the more experience and familiarity the examinees gained, the better their results. Some studies have found that a lack of computer familiarity may lead to lower performance (Chan et al., 2018; Lee, 1986), findings consistent with the present study. Nonetheless, the effect of computer familiarity, identified by some studies, was negligible, as no relationship was found between it and test performance (Burgmanis, 2023; Taylor et al., 1999). Despite miscellaneous findings in other research, all emphasized the necessity of tutorials or instructions for computerized tests. Like Yeom and Jun's (2020) study, our research concluded that the test-taker's unfamiliarity with CALT could affect the students' preferences. However, they did not find any significant difference in their performance. New computer literacy and standardized testing processes should equip test-takers with superior technology and testing insight by providing opportunities and enabling them to develop the required computer self-efficacy. Finally, like Brown's (2019) study, our research aimed to demonstrate how transitioning to computerized testing presents numerous opportunities that require careful consideration before engaging in the development process.

Unlike the Zhang and Razak's (2026) findings, computer anxiety in this study was attributed to a more general anxiety, causing emotional disorder. As such, the anxiety felt by the testees at PNU was not attributed to computerphobia, but rather to other fears, such as being unable to

access the exam portal, failing to maintain data during the exam, closing the program, and submitting answers successfully. As various factors related to examinees, facilitating conditions, and result demonstrability can enhance computer anxiety due to lack of experience and unfamiliarity with computers (Zhang & Razak, 2026), it can affect test performance accordingly. Although computer anxiety is mixed up with test anxiety, if the test is delivered through computers, the impact of both on the test results deteriorates. However, complicated tests are affected only by test and computer anxiety, whereas multiple-choice tests place lower cognitive demands; thus, test anxiety does not significantly affect test performance. These types of questions do not require higher-level processing, reasoning, or deductive thinking to complete more than 2 tasks simultaneously. A negative effect on testees' performance was ascertained due to high levels of CA on complex CBTs (Nicholson et al., 2003). As most English tests at PNU require testees to memorize and recall receptive knowledge, they do not cognitively exploit cognitive capacity or WM.

Computer literacy always assumes that test takers' computer use is either positively or negatively influenced. To ensure that the technology employed in language testing is not biased in favor of one group or detrimental to particular subgroups of examinees (Roever & McNamara, 2006), test-takers must possess equal knowledge and basic familiarity with computers. Some research found no relationship between computer literacy and students' performance on computers (Taylor et al., 1999), which was consistent with this study. Reviewing 105 studies on digital literacy, familiarity, and their relation to construct-irrelevant digital demands, Zhang and Razak (2026) found that unfamiliarity with the system can enhance anxiety during online assessment can reduce the performance. They highlighted the crucial impact of platform design, preparation, and reporting on language ability. Likewise, Chapelle and Voss (2016) stressed that test takers should be prepared for the CALT exam before the exam to avoid the unwanted disadvantage of lacking technology literacy. Examinees with less computer experience had little trust in computerized tests; therefore, universities should offer preparatory workshops to build examinees' confidence when encountering computer problems. By doing so, PNU can ensure that fairness and claims are upheld, as the test methods do not affect the results (Dadey et al., 2018).

The present study was closely aligned with Habibi's (2017) investigation, which emphasized that PNU administrators must use high-quality equipment for the test method and train technical staff. Both studies identified amenability as an essential factor in CBT. For instance, technical personnel need to be available to fix problems so that testees feel secure during the test and can concentrate fully on the exam itself. Any prior negative experience can cause a test taker to develop a negative perspective towards CALT, thereby affecting their full attention and, consequently, their performance on the test. However, professional programming can mitigate most malfunctions, enhancing the quality of assessments or tests. By doing so, test-takers feel thoroughly confident and secure when using computers on the tests. As a result, their performance is automatically enhanced. The development of new devices, such as tablets, smartphones, and PCs, along with ongoing improvement in online assessments on genuine tutorial websites, has facilitated language learning, teaching (Golonka et al., 2014), and testing. Although computers have presented some issues, such as maintenance burdens and facility operating costs, they have also brought opportunities in language testing at PNU.

The findings of the study were compatible with that of Olufemi and Oluwatayo (2014) in that the existence of computer literacy and the absence of CA enhance test performance and self-efficacy (Lee & Huang, 2014) at the same time, computer literacy reduces CA; hence the fair evaluation of performance in CALT requires the reduction of CA and enhancement of computer familiarity (Clariana & Wallace, 2002; Taylor et al., 1999). Since computers facilitate teaching foreign languages (Golonka et al., 2014) and testing (Chapelle & Voss, 2016), preliminary instruction is required to help the students employ computers for language learning and testing at their own pace in the Iranian context. Likewise, as in the study by Ebadi and Goodarzi (2017), this research also found positive attitudes toward CALL. PNU courses enable students to study at their own pace while employed, making them more willing to learn and utilize self-access activities and distance learning.

Exploring the experience of automatic grading, the majority of testees associated it with saving time and energy and with relieving anxiety. This is because test takers' performance on computerized tests can be evaluated immediately by both human raters and computers (Suvorov & Hegelheimer, 2014). The participants explained that they needed to study only a series of past tests that they could pass, a practice known as teaching to the test or Washback effects (Alderson & Wall, 1993) or impact (Bachman et al., 2010). However, a few of them also confessed that they had studied the footnotes and reviewed the books four times to memorize the materials. To maintain validity, they must be replaced with a comprehensive alternative assessment (Stobart & Gipps, 2010). They adapted dynamic assessment as a mediation beyond the original notion of psychological tools (Lantolf et al., 2015) to enhance teaching and testing methods (Daniels & Gierl, 2017) by integrating learning and teaching.

As CALT experience is gradually gained through exams, the examinees' positive perspectives towards CALT will evolve steadily. A compelling justification for this claim is that success or failure in examinations seems closely tied to the time, effort, and expenditures Iranian students invest. Since the results revealed that the examinees' positive perspectives increased with CALT experience but not with higher academic degrees, students should be exposed to more computerized exams to become more familiar with them. Additionally, a degree did not moderate the relationship between CALT experience and examinees' perspectives; therefore, the computer familiarity or CALT experience of well-educated examinees should not be taken for granted. On the other hand, more experienced individuals are more inclined toward CALT, and gaining CALT experience helps maintain a positive perspective toward it, resulting in better test performance. Notwithstanding, a higher degree did not affect testees' perceptions regarding CALT, i.e., graduates and undergraduates did not differ in their inclination towards CALT

Conclusion

The study found that although PNU's policy is to reduce energy, effort, and time consumption, computer maintenance—the practice of keeping computers and their components in good working order—can cause problems. However, distribution expenses and human resources are required to prepare and distribute the test sheets, which necessitate special funds and effort. Overall, these findings can provide constructive guidance for test developers and universities to focus on more standardized tests, thereby enhancing the professional testing system and

delivery. Namely, designing a more professional test format and delivery system can enhance test reliability and validity, resulting in a more consistent test. Universities also need to recognize that, if there are systemic test problems, there is a higher risk of inconsistency across test administrations, leading to measurement errors that are contrary to the ultimate goal of reliability. Likewise, it would be hard to determine how much of the test performance is due to measurement error or to factors other than underlying language ability, which relates to construct validity. Hence, this study can provide the educational system with insights to enhance test delivery and avoid inconsistencies.

Despite these contributions, this study has several limitations that warrant mention. While this study provided a detailed view of computerized testing, the findings may not be applicable outside the PNU setting. Future research could replicate this study using different subject materials in various contexts to determine whether the same results are achieved. Second, considering variables from the participants' perspective may not be a reliable criterion for examining the test's bias; therefore, the results may be biased, as participants might hold inaccurate views of the testing system and may not fully align with their actual perspectives. Using quantitative methods, such as measuring test results and comparing them with actual performance in a reliable testing system, would have provided richer insights into the different concepts in practice. Finally, the unavailability of the participants' actual scores led the researchers to evaluate the testing system from the examinees' perspective.

The results suggest a need for a more comprehensive approach to computerized tests that also addresses the emotional and cognitive aspects of computer-based testing in the university context. Moreover, the correlation between test anxiety and computer-based anxiety can also be examined to ensure the correlation between these two variables. Another study can be conducted to investigate the effect of test format and examinees' performance in two different modes of delivery, and to determine which group performs far better.

Bio-data

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest: The authors declare that they have no competing interests.

References

- Akdemir, O., & Oguz, A. (2008). Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education*, 51(3), 1198-1204. <https://doi.org/10.1016/j.compedu.2007.11.007>
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Alrob, M., Abu, M., Asad, N. A., & Daqar, M. A. (2019). Attitudes toward and Implications of the Computer-Based Exams at Arab American University of Palestine. *Journal of Education and Learning*, 8(1), 196-205. <https://10.5539/jel.v8n1p196>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press Oxford.
- Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4), 449-465.
- Bahari, A. (2021). Computer-assisted language proficiency assessment tools and strategies. *Open Learning: The Journal of Open, Distance and e-Learning*, 36(1), 61-87. <https://doi.org/10.1080/02680513.2020.1726738>
- Balogun, A. G., & Olanrewaju, A. S. (2016). Role of computer self-efficacy and gender in computer-based test anxiety among undergraduates in Nigeria. *Psychological Thought*, 9(1), 58-66. <https://doi.org/10.5964/psyct.v9i1.160>
- Brown, G. T. (2019). Technologies and infrastructure: costs and obstacles in developing large-scale computer-based testing. *Education Inquiry*, 10(1), 4-20. <https://doi.org/10.1080/20004508.2018.1529528>
- Bugbee, A. C. Jr. (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>
- Burgmanis, Ģ., Mikīte, M., France, I., & Namsone, D. (2023). Development of computer-based diagnostic assessment system: case study of the equivalence of paper-and-pencil and computer-based testing. In *Society Integration Education Proceedings of the International Scientific Conference* (Vol. 1, pp. 613-622).
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593-602.
- Chan, S., Bax, S., & Weir, C. (2018). Researching the comparability of paper-based and computer-based delivery in a high-stakes writing test. *Assessing Writing*, 36, 32-48. <https://doi.org/10.1016/j.asw.2018.03.008>
- Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press.
- Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology* (Vol. 7). John Benjamins Publishing.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Ernst Klett Sprachen.
- Chapelle, C. A., & Sauro, S. (2017). *Introduction to the handbook of technology and second language teaching and learning*. Wiley Online Library.
- Chapelle, C. A., & Voss, E. (2016). Twenty years of technology and language assessment in Language Learning and technology. *Language Learning & Technology*, 20(2), 116-128. <https://www.learntechlib.org/p/176101/>.

- Dadey, N., Lyons, S., & DePascale, C. (2018). The comparability of scores from different digital devices: A literature review and synthesis with recommendations for practice. *Applied Measurement in Education*, 31(1), 30. <https://doi.org/10.1080/08957347.2017.1391262>
- Daniels, H. (2015). Mediation: An expansion of the socio-cultural gaze. *History of the Human Sciences*, 28(2), 34-50. <https://doi.org/10.1177%2F0952695114559994>
- Daniels, L. M., & Gierl, M. J. (2017). The impact of immediate test score reporting on university students' achievement emotions in the context of computer-based multiple-choice exams. *Learning and Instruction*, 52, 27-35. <https://doi.org/10.1016/j.learninstruc.2017.04.001>
- Dewi, A. T. A., Rofik, M., Holidi, M., Najiburahman, N., Baharun, H., Zaini, A. W., & Sanjani, M. A. F. (2025). CBT exam implementation guidance for students in facing digital-based assessments. *Communaautaire: Journal of Community Service*, 4(3), 297-312.
- Douglas, D. (2012). Technology and language testing. *The encyclopedia of applied linguistics*. <https://doi.org/10.1002/9781405198431.wbeal1182>
- Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning*, 32(5-6), 527-555. <https://doi.org/10.1080/09588221.2018.1527362>
- Ebadi, S., & Bashir, S. (2020). An exploration into EFL learners' writing skills via mobile-based dynamic assessment. *Education and Information Technologies*, 1-22.
- Gipps, C., & Stobart, G. (2010). Alternative assessment. In *International Encyclopedia of Education* (3rd ed.). Elsevier.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70-105.
- Habibi, H. (2017). Pathology of electronic tests in Payame Noor University. *Educational Measurement and Evaluation Studies*, 6(16), 43-66.
- Heigham, J., & Croker, R. (2009). *Qualitative research in applied linguistics: A practical introduction*. Springer.
- Hinton, P., Brownlow, C., & McMurray, I. (2004). I., & Cozens, B. (2004). *SPSS explained*.
- Hosseini, M., Abidin, M. J. Z., & Baghdarnia, M. (2014). Comparability of test results of computer based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran. *Procedia-Social and Behavioral Sciences*, 98, 659-667. <https://doi.org/10.1016/j.sbspro.2014.03.465>
- Jimoh, R., Shittu, A. J. K., & Kawu, Y. K. (2012). Students' perception of computer based test (CBT) for examining undergraduate chemistry courses. *Journal of Emerging Trends in Computing and Information Sciences*, 3(2), 125-134. <https://10.14738/abr.46.2325>
- Jin, Y., & Yan, M. (2017). Computer literacy and the construct validity of a high-stakes computer-based writing assessment. *Language Assessment Quarterly*, 14(2), 101-119. <https://doi.org/10.1080/15434303.2016.1261293>
- Khoshsima, H., & Toroujeni, S. M. H. (2017). Computer-Based Testing: Score Equivalence and Testing Administration Mode Preference in a Comparative Evaluation Study. *International Journal of Emerging Technologies in Learning*, 12(10).
- Khoshsima, H., Hashemi Toroujeni, S. M., Thompson, N., & Reza Ebrahimi, M. (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.
- Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2015). Sociocultural theory and second language development. *Theories in second language acquisition: An introduction*, 1, 207-226.

- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R*: Routledge.
- Lee, J. A. (1986). The effects of past computer experience on computerised aptitude test performance. *Educational and Psychological Measurement*, 46(3), 727-733. <https://doi.org/10.1177%2F0013164486463030>
- Lee, C. L., & Huang, M. K. (2014). The influence of computer literacy and computer anxiety on computer self-efficacy: the moderating effect of gender. *Cyberpsychology, Behavior, and Social Networking*, 17(3), 172-180. <https://doi.org/10.1089/cyber.2012.0029>
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Nicholson, J. A., Hardin, A. M., & Nicholson, D. B. (2003). *Test performance and the medium: Unearthing differences that make a difference* [Conference presentation]. Americas Conference on Information Systems.
- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal*, 93, 836-847. <https://doi.org/10.1111/j.1540-4781.2009.00976.x>
- Olufemi, O. A., & Oluwatayo, O. J. (2014). Computer anxiety and computer knowledge as determinants of candidates' performance in computer-based test in Nigeria. *Journal of Education, Society and Behavioural Science*, 495-507. <https://doi.org/10.9734/BJESBS/2014/6632>
- Pallant, J. (2011). Survival manual. *A step by step guide to data analysis using SPSS*.
- Powell, A. L. (2013). Computer anxiety: Comparison of research from the 1990s and 2000s. *Computers in Human Behavior*, 29(6), 2337-2381. <https://doi.org/10.1016/j.chb.2013.05.012>
- Richards, L. (2014). *Handling qualitative data: A practical guide* (3rd ed.). SAGE Publications.
- Roever, C., & McNamara, T. (2006). *Language testing: The social dimension*. Blackwell Publishing
- Schulenberg, S. E., & Melton, A. M. (2008). The computer aversion, attitudes, and familiarity index (CAAFI): A validity study. *Computers in Human Behavior*, 24(6), 2620-2638. <https://doi.org/10.1016/j.chb.2008.03.002>
- Suvorov, R., & Hegelheimer, V. (2014). Computer-assisted language testing (w) Kunnan AJ (red.), *The Companion to Language Assessment* (Vol. 2, Chap. 36). In: London: John Wiley & Sons.
- Taylor, C., Kirsch, I., Jamieson, J., & Eignor, D. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning*, 49(2), 219-274. <https://doi.org/10.1111/0023-8333.00088>
- Tella, A., & Bashorun, M. (2012). Attitude of undergraduate students towards computer-based test (cbt): a case study of the university of Ilorin, Nigeria. *International Journal of Information and Communication Technology Education (IJICTE)*, 8(2), 33-45. <https://doi.org/10.4018/jicte.2012040103>
- Voss, E. (2018). Technology and Assessment. *The TESOL Encyclopedia of English Language Teaching*, 1-7.
- Winke, P. M., & Isbell, D. R. (2018). Construct of listening. *The TESOL Encyclopedia of English Language Teaching*, 1-6. <https://doi.org/10.1002/9781118784235.eelt0618>
- Wise, S. L. (2019). Controlling construct-irrelevant factors through computer-based testing: disengagement, anxiety, & cheating. *Education Inquiry*, 10(1), 21-33. <https://doi.org/10.1080/20004508.2018.1490127>
- Yeom, S., & Jun, H. (2020). Young Korean EFL Learners' Reading and Test-Taking Strategies in a Paper and a Computer-Based Reading Comprehension Tests. *Language Assessment Quarterly*, 1-18. <https://doi.org/10.1080/15434303.2020.1731753>

Zhang, J. M., & Abdul Razak, R. B. (2026). Literature Review: Analysing the Factors Affecting College Students' Digital Language Tests. *Africa Education Review*, 1-36.

Appendix

Perspectives of the English Language Students and Instructors toward Computer-Assisted Language Testing This questionnaire aims to discover Iranian students' perspectives toward computer-assisted language testing. There is no right or wrong answer on this scale. Please, mark the answer that represents your stance toward each item in the scale.

First & Last name (optional).....

What is your gender? Female ☐ Male ☐

How old are you?.....

How many times have you ever taken English exams by computers?

Never ☐ Less than five times ☐ More than five times ☐

What is your major? English literature ☐ English translation ☐ TEFL ☐

I am a student ☐ I am an English teacher ☐

Email address (optional).....

Strongly agree (SA), Agree (AG), Uncertain (UC), Strongly disagree (SD), Disagree (DA)

Questionnaire	SA	AG	UC	SD	DA
1. Computers should be used in language assessment.					
2. If I had a chance to select the test mode, I would prefer to take the exam through computers.					
3. CBTs are safer than Paper-and-pencil.					
4. Computers should be applied at all Universities for testing languages more than they are used now.					
5. Computer-based test (CBT) ¹ can assess higher-level thinking skills. ²					
6. Test administrators should be skillful enough in using technology.					
7. I believe that technology facilitates testing.					
8. It is easy for me to use computers for exams.					
9. I have enough skills to use computers on the exams.					
10. Using computers on exams saves time.					
11. I am not afraid to engage with computers on exams.					
12. Multiple choice questions are more appropriate than yes/no questions in the CBT.					
13. I get more tired taking the exams by paper-and-pencil than taking the exams by computers.					
14. Gaining the results immediately after the exams by computers in the exams reduce my stress.					
15. CBT is easy to read on the screen and takes less time than reading on paper-and-pencil tests.					
16. The CBT frees me from tedious reading, thereby concentrating on more complicated tasks.					

¹ Taking the test through computers

² It distinguishes critical thinking skills from low-order learning outcomes, such as those attained by rote memorization including synthesizing, analyzing, reasoning, comprehending, application, and evaluation.

17. Cheating is easier in paper exams than CBTs.					
18. CBT is more efficient and more comfortable than paper-and-pencil testing.					
19. It is worth employing computers for language testing despite some unexpected problems.					
20. The CBTs are free of knowledge & scientific mistakes.					
21. The CBT environment is more conducive for an examination compared to paper exams.					
22. Technical staff helps students adequately in their technical inquiries.					
23. The CBTs are also suitable for students with special needs.					
24. CBTs well reflect the course content at PNU.					
25. Two exams for each course (midterm and final) are appropriate in the new system.					
26. The CBT requires technical support ³					
27. The CBT can be completed within the set time.					
28. My CBT results are better than my paper exam results.					
29. I rely on guesswork ⁴ to answer the questions on paper exams more than I do on CBT.					
30. I believe that I can take all my exams with computers.					
31. The CBTs are free of linguistic mistakes.					
32. The results of CBTs are objective. ⁵					
33. The CBT instructions are clear and easy to understand.					
34. The time allocated for exams is adequate.					
35. The CBTs contribute to improving the learning process.					
36. The CBT questions reflect the content learned in the classroom.					
37. CBTs raise the level of anxiety.					
38. Essay ⁶ questions can be included in CBTs.					
39. CBTs are adequately aligned with our course content.					
40. CBT questions can be leaked.					

Please write whatever you think may help us improve computer-based tests.

³ For any possible problems may occur during the exam.

⁴ Answering the choices by guessing

⁵ In multiple choice type

⁶ Essay: questions require typing like writing