



Computer-Based English Competency Assessment for Scholarship Selection: Challenges, Strategies, and Implementation in the Ministry of Finance

Tes Kemampuan Bahasa Inggris Berbasis Komputer untuk Seleksi Beasiswa: Tantangan, Strategi, dan Implementasi pada Kementerian Keuangan

Muhammad Yoga Prabowo¹, Sarah Rahmadian²

Melbourne Graduate School of Education, The University of Melbourne, Carlton Victoria, Australia¹

Human Resource Development Education and Training Center, Financial Education and Training Agency, Ministry of Finance, DKI Jakarta, Indonesia²

muhammadyoga@student.unimelb.edu.au

ARTICLE INFO

Keywords:

computer-based testing, English language assessment, scholarship selection, Ministry of Finance, Ministerial Scholarship

ABSTRACT

The need for a safe and versatile testing system during the COVID-19 pandemic for scholarship selection prompted the Financial Education and Training Agency (FETA), an echelon-1 unit under the Ministry of Finance, to design the *Aplikasi Ujian Online* computer-based test platform. This paper focuses on exploring challenges in developing and administering English competency assessment through the computer-based platform and the strategies to overcome them. The research used the descriptive method with direct observation and questionnaire for data collection. Three receptive skills (listening, structure, and reading comprehension) were measured on 337 applicants of the Ministerial Scholarship offered by the Ministry of Finance in 2021. A post-test questionnaire was administered to investigate test takers' perceptions. The findings showed that loading time and login difficulty were the most common issues. However, preformulated risk-mitigation strategies were found to be effective in resolving the issues. Overall, the questionnaire result showed a significantly positive perception of the computer-based English competency test.

INFO ARTIKEL

Kata kunci:

ujian berbasis komputer, penilaian bahasa Inggris, seleksi beasiswa, Kementerian Keuangan, Ministerial Scholarship

ABSTRAK

Kebutuhan suatu sistem tes yang aman dan serbaguna untuk seleksi beasiswa mendorong Badan Pendidikan dan Pelatihan Keuangan (BPPK), sebuah unit eselon 1 di bawah Kementerian Keuangan, untuk merancang sistem ujian berbasis komputer bernama *Aplikasi Ujian Online*. Penelitian ini berfokus untuk mendalami tantangan dalam mengembangkan dan menyelenggarakan penilaian kemampuan bahasa Inggris dan strategi untuk mengatasinya. Metode yang digunakan dalam penelitian ini adalah metode deskriptif dengan teknik pengumpulan data berupa pengamatan langsung dan penyebaran kuesioner. Tiga kecakapan reseptif (listening, structure, dan reading comprehension) diujikan pada 337 pendaftar beasiswa Ministerial Scholarship yang ditawarkan oleh Kementerian Keuangan pada tahun 2021. Kuesioner pascaujian disebarkan untuk mengetahui persepsi peserta ujian. Hasil penelitian menunjukkan waktu muat dan kesulitan login merupakan masalah paling umum yang dihadapi oleh peserta tes. Namun demikian, strategi mitigasi risiko yang telah disusun sebelumnya dirasa efektif untuk menyelesaikan masalah yang terjadi. Secara keseluruhan, hasil kuesioner menunjukkan peserta tes memiliki persepsi yang sangat positif mengenai uji kemampuan bahasa Inggris berbasis komputer.

Introduction

The Ministerial Scholarship aims to support the Ministry of Finance's competency development to achieve its vision, mission, and strategic goals through sending future leaders and best talents to study abroad. The program, operated by the Division of Scholarship Management of the Financial Education and Training Agency, offers funding for the Ministry of Finance's employees to take master's and doctoral degrees at the world's top 30 universities. The selected awardees should follow a competitive selection process comprising academic selection, psychological assessment, and panel interview according to the Decree of the Head of the Financial Education and Training Agency Number 2 of 2020 concerning Degree Program Management Technical Policies in the Ministry of Finance.

Two types of tests are performed in academic selection: the Academic Potential Test and the English Competency Test. In the past, the tests were administered through the paper-and-pencil method. The production and distribution of test booklets were a time-consuming and labor-intensive process. Test booklets were printed, bound, and sealed into secured envelopes before being packed into luggage. Representatives of the central committee would bring the test booklets to each test site through air travel and then return the filled answer sheets after the test. Once the central committee received the answer sheets, the evaluation team would manually unpack and scan the sheets for scoring.

Since this paper-based test (PBT) mode involves a substantial amount of physical contact during the whole process, computer-based testing (CBT) was deemed a safer and more suitable option from the standpoint of healthcare. Restrictions of travel and public activities during the pandemic also made tests in the conventional method unfeasible. In addition, computer-based testing offers other advantages, such as expedient delivery, digital record-keeping options, and automatic scoring possibilities (Winke and Isbell, 2017).

Prior to this selection, the Ministry of Finance had used computer-based testing on several occasions. *Aplikasi Ujian Online* is a web-based test platform developed by the Division of Information Technology and Communication of the Financial Education and Training Agency to accommodate various testing needs in the Ministry of Finance. However, it was the first time that computer-based testing had been used to select scholarship candidates in the ministry.

The Degree Program Management Technical Policies have made a distinction between internal scholarship selection administered independently by the Ministry of Finance, referred to as the Ministerial Scholarship, and external scholarship programs, referred to as the Institutional Scholarship. This study is mainly focused on the Ministerial Scholarship.

Scholarship selection is considered a high-stake test with profound consequences. As a fully funded scholarship program, the Ministerial Scholarship offers awards totaling billions of rupiahs for studying abroad. It covers tuition fees, living allowance, international travel expenses, health insurance, and other study-related expenses. Consequently, the scholarship is highly desirable and competitive. Every year, up to a thousand employees apply for this scholarship, despite the fact that the number of accepted applicants is limited to 40-80.

Maintaining fairness in all aspects of the process is a crucial issue. A challenging problem that arises in this domain is cheating. Test takers may attempt to cheat in order to gain an advantage. This problem has attracted the attention of many researchers. A study by Alessio and Messinger (2021) on university faculty members and students revealed that 81% of faculty members agreed that cheating in online tests is easier, while 83% of students had the same perception. Additionally, Chirumamilla, Sindre, and Nguyen-Duc (2020) identified six different cheating practices: impersonation, forbidden aids, peeking, peer collaboration, outside assistance, and student-staff collusion. They found that both students and teachers thought cheating was easier on CBT, particularly those delivered on students' own devices. On a positive note, CBT also allows for easier implementation of strategies to counter dishonest behaviors.

Therefore, the organizing committee should carefully consider many aspects and all possible outcomes before deciding on the test design and implementation. Because the test may take place in more than one site, the committee must create guidelines to ensure that the facilities, supervision, and administration are consistent across all locations. Furthermore, the test must be able to accommodate a wide range of test takers.

English language proficiency is a widely used requirement in scholarship application and university admission. While other scholarship providers accept scores from established standardized English language proficiency tests, the Ministerial Scholarship is one among very few scholarship providers in Indonesia that conduct their own assessments. The Division of Integrated Test Management, established in 2011, is an echelon-3 unit under the Financial Education Training and Agency responsible for planning, creating, and developing test items for Academic Potential Test, English Competency Test, and psychological testing in the Ministry of Finance.

While the item type of the computer-based English Competency Test is identical and presented linearly as its paper-based counterpart, it differs in some respects from other tests accommodated by *Aplikasi Ujian Online*. The Listening Comprehension section, in particular, poses a major challenge since the audio is loaded into each test taker's computer instead of being played on a centralized sound system. This configuration requires a headset to listen individually, a feature to embed audio, and more bandwidth than text-based questions. The Listening and Reading Comprehension sections also contained testlets (a group of test items that are administered together) which might be challenging to display on the screen.

Recent studies have shown that test takers in high school and higher education settings had a generally positive perception of the computer-based test worldwide. Khoshshima and Toroujeni (2017) observed that 75% of Iranian graduate students preferred CBT to PBT in the General English assessment. In Indonesia, Sulistiyono, Suyata, and Rahayu (2018) found that teachers, headmasters, and students generally perceived CBT for *Ujian Nasional Berbasis Komputer* (Computer-Based National Examination) more favorable than the conventional counterpart. Similarly, Sulastris and Puspawati (2019) reported that university students who had experienced CBT during their high school study reported positive feelings in using CBT for standardized testing.

Outside the domain of educational institutions, the usage of CBT has rarely been studied directly. Only a few studies have been undertaken on selections conducted by government agencies. Since they depend on the state budget for financing, it is critical to look into the accountability of their programs. With this in mind, we therefore analyzed the implementation of computer-based English competency assessment for the scholarship selection in the Ministry of Finance. This present study was conducted to identify and describe 1) the challenges faced before and during the implementation of CBT, 2) the strategies used by the committee to overcome the challenges, 3) the perceptions of the test takers on the delivery of the CBT.

Literature Review

CBT promises advantages over traditional testing methods. However, it is critical to analyze not only the positive benefits but also potential negative implications. While a wide body of research has attempted to classify the challenges of CBT, most are limited in scope and often overlook particular groups of participants, such as test takers with disabilities. In this study, we use a comprehensive categorization proposed by Thurlow et al. (2010), which identified four categories or "themes" to organize the challenges of CBT.

Table I Categories of CBT Challenges

Categories	Description
Economic	Factors that have cost implications
System implementation	Logistical, test security, and other factors that affect the capacity of education agencies to implement CBT
Test administration/design	Factors that affect how test takers perceive and interact with the test
Accessibility	Factors that affect how accessible a test is to a wide range of test takers

(Source: Thurlow et al., 2010)

To overcome the challenges posed by the CBT, policymakers should devise strategies prior to implementing the test. A recent study by Sulastrri and Puspitawati (2019) suggests that strategies to face challenges in CBT implementation can be classified into two major themes: technical and non-technical strategies. Technical strategies are concerned with physical facilities, whereas non-technical strategies are concerned with human resources and other intangible aspects.

Proctoring is one way to prevent cheat and fraud during computer-based tests. Proctor can be a person or software. Proctoring software comes with various capabilities, including screen lockdown, disabling screen capture, and camera recording. Medina and Castleberry (2016) suggested that when making the transition from paper-based to computer-based tests, the test committee must create a detailed proctoring policy to standardize procedures and ensure accountability. Therefore, proctoring policy should be explored as an important factor.

Method

The methodology employed for this study is descriptive research. Nassaji (2015) argues that the goal of descriptive research is to describe a phenomenon and its characteristics. Descriptive research is primarily focused on what rather than how or why something occurred. This methodology is suitable to provide systematic information that this study seeks to present.

This current study was carried out on the academic selection process of the Ministerial Scholarship selection, which took place on 6–8 April 2021. The data were collected through direct observation and questionnaire. Direct observation is when researchers watch interactions, processes, or behaviors as they occur (CDC, 2018). The observation was mainly focused on the organizing committee during the planning, implementation, and evaluation stages.

After the test, an online questionnaire was distributed to the test takers to investigate their perception of the computer-based test. The instrument used five-point scale items (1 = poor, 2 = fair, 3 = average, 4 = good, 5 = excellent) and open-ended questions to examine the respondents' reasons. The collected data were analyzed using frequencies, percentages, and averages to gain a deeper understanding of test takers' opinions and attitudes.

The questionnaires were completed by 336 people out of 337 who took the test. The number of male respondents (n=78%) outnumbers the number of female participants (n=22%) by a substantial margin. All of the respondents were between the ages of 26 and 42. The average age was 33.45 years with a standard deviation of 3.83. When missing values in the questionnaire's response were found, they were assigned the value of N/A.

Results and Discussion

Challenges and Strategies

The implementation of CBT for scholarship selection in the Ministry of Finance in general was met with some challenges. The challenges are presented in four categories: economic, system implementation,

test administration/design, and accessibility. The strategies formulated to overcome the challenges are presented in the subsequent column.

Table II Economic Challenges

Challenges	Strategies
Limited facilities (rooms, PCs, headsets)	Dividing the test into several sessions; requesting test takers to bring their own headsets
Platform development costs	Developing independently (using in-house programmers)

The economic challenges of implementing CBT are mostly related to the hardware (facilities) and the software (testing platform). Initially, the test committee considered three alternative schemes of test location. These schemes affected how the facilities were arranged. First, conducting the test at regional offices of the Financial Education and Training Agency. The Financial Education and Training Agency has a polytechnic and eleven regional campuses across the country. These regional campuses formed regional committees to help the central committee organize the test. Their responsibilities included arranging testing rooms, assigning proctors, and coordinating test preparations and delivery with the central committee. For small-scale tests, these campuses have sufficient computers/laptops for CBT. In some locations with a large number of test takers, the test was conducted on more than one day with two sessions each. The morning session was designated for the Academic Potential Test and the afternoon session was for the English Competency Test. Additionally, this arrangement also helped the committee preventing COVID-19 transmission through the mass gathering.

Second, conducting the test at test takers' respective offices or offices owned by their directorate generals. Since all test takers were the Ministry of Finance employees, it was deemed possible to arrange this scheme. This scheme, however, is more difficult to control. While it is possible to enlist the help of the internal compliance subdivision in each office for proctoring, there is a substantial technical challenge since different offices may or may not have supporting facilities such as an emergency power supply system.

Third, conducting the test at test takers' home. This means that the test is conducted fully online using test takers' own devices. The most serious drawback is that no proctors monitor the test takers on site. To guarantee that they are not cheating, the test takers must provide a second device connected to a cloud meeting system where the central committee can monitor them. Another problem was unequal access to technology. Some test takers may find it difficult to provide a second device. Furthermore, the internet connection speeds in different locations might vary, making it difficult for some test takers to access the testing platform.

After carefully considering possible schemes, the central committee decided to implement the first scheme. The test venues were spread throughout 11 cities: Jakarta, Medan, Pekanbaru, Palembang, Cimahi, Yogyakarta, Malang, Denpasar, Pontianak, Pekanbaru, Balikpapan, Makassar, and Manado. For a test of this scale (341 registered participants), there was no additional cost incurred for procuring additional computers. Another challenge was the availability of headsets required for the Listening section. The committee then instructed the test takers to bring their own headsets. This also could reduce virus transmission from using shared headsets.

Compared to PBT, CBT systems in general are initially more costly to develop. Therefore, instead of hiring external developers, the Financial Education and Training Agency developed the CBT system independently. The Division of Information Technology and Communication played an important role in developing the CBT platform in the form of *Aplikasi Ujian Online*. While the initial development of the CBT platform itself predated the need for online scholarship selection, further development was required to suit the platform with particular testing needs.

In contrast to the notion that CBT systems are more expensive than PBT, the implementation of CBT in the Ministry of Finance was in fact more cost-effective since it reduced the expenditure for producing and transporting the test booklets. CBT implementation for the scholarship selection in 2021 has reduced production costs and eliminated land transport, airfare, and hotel accommodation expenses for locations outside Jakarta.

Table III System Implementation Challenges

Challenges	Strategies
“Generic” platform design	Tailoring platform to suit testing needs
Cheating possibility	Using human and software proctoring; using multiple item sets; using face verification; using access tokens
Item confidentiality	Assigning different access rights based on roles; disabling copy-paste and screen capture
Network	Requesting additional bandwidth on test days
Power	Using on site emergency power generators

Aplikasi Ujian Online was designed as a general test management platform to accommodate various types of tests. As a result, some features which are unique to the English language test might not be present or require adjustments. The developer of the platform had included the capability to embed audio. However, the feature had never been validated before and therefore must be proven to be able to endure the stress of large-scale access. On 18–22 February 2021, a pilot test was conducted on 176 participants to gain information for platform improvement. The pilot test was delivered in two modes: on participants’ devices at home and test centers. The result showed that the test on participants’ devices was more vulnerable to technical issues, particularly during the Listening section. While the test conducted at test centers showed a better result, there were reports of considerable loading time before someone could start the Listening section. This result prompted the platform developer to find solutions, such as using catching techniques on routers to improve speed and stability. Additionally, the committee received a large number of suggestions for improving the user experience, which were duly followed up.

In safeguarding the test from cheating, the committee implemented rigorous protective measures. In addition to human proctors, a software proctoring system was used to prevent test takers from seeking answers on the internet or running other applications. This was accomplished by using the Safe Exam Browser, a security browser used in conjunction with the online testing platform. In *Aplikasi Ujian Online*, the test administrator can set the test to be run on Safe Exam Browser only. When using the Safe Exam Browser, the test is displayed in full-screen mode and access to other pages, software, or certain keys is restricted (similar to a lockdown). As a result, test takers cannot copy or take screenshots of the questions. Additionally, they cannot ask for external assistance through communication software.

However, the proctoring software cannot prevent someone from peeking at the answers of other test takers physically. To prevent this type of cheating, the committee assigned a pattern of seating arrangement and multiple sets of items so that adjacent test takers would get different questions. This was done not by merely shuffling the order of questions or choices but by using different parallel forms. A form is a version of a test written to the same specifications and developed to measure the same construct (American Educational Research Association [AERA] et al., 2014).

Another type of cheating is impersonation. Test takers might hire someone else, often referred to as “jockey”, to take the exam for them. This is often seen as a lucrative business for syndicates that offer to help someone pass entrance exams into universities or public service careers in dishonest ways. Frankl, Scharner, and Zebedin (2012) suggested that this kind of cheating can only be prevented by organizational measures like face-to-face verification. Test takers were asked to fill out an online attendance form on their mobile phones before each test to thwart impersonators in the scholarship selection. They also took

a close-up photograph of their faces in the same form. These pictures were compared with the test takers' official portraits retrieved from the registration portal and HR information system.

In an online examination, a “jockey” may gain access to the CBT platform from outside the test center if he knows the test taker's password and the platform's internet address. To prevent unauthorized access, the committee implemented an additional security measure in the form of access tokens. An access token of a section was different from that of others, without which one could not begin the test. This helped ensure that even if the test takers gave their passwords away, no one else could have unauthorized access to the test.

Item confidentiality was also a major concern for the Division of Integrated Test Management since the items were difficult to construct and quite expensive if it is purchased from external parties. If the test items leaked, the fairness of the test would be compromised. Therefore, different access rights were applied to the *Aplikasi Ujian Online* administrator portal. There were separate accounts with different access rights for the Division of Integrated Test Management administrators and regional committees' proctors who supervised the test onsite. Unlike the administrators, the proctors could not access the test questions.

Network and power are two necessities to deliver a computer-based test. The committee requested the Center for Financial Information System and Technology, an echelon-2 unit within the Ministry of Finance with the authority to control the ministry-wide internet access, to increase the bandwidth of the test centers. In addition, emergency power generators were also prepared on all test centers to mitigate the risk of power outages. There was no reported power outage throughout the test. However, an unexpected network problem occurred in one location when a test taker inadvertently kicked the network switch, causing it to turn off. As the test was run on the client-side, it did not affect the test takers directly. The connection to the server was restored as soon as the network was reestablished.

Table IV Test Administration/Design Challenges

Challenges	Strategies
Test taker readiness	Offering dummy test; conducting a pilot test to gauge user experience
User-friendliness	Designing a test platform with a similar design to a familiar e-learning platform
Errors during tests/technical issues	Special circumstances guideline; troubleshooting practice
Proctors	Providing training for proctors
Health concern	Temperature scanners; PCR/swab test result; vaccination certificates; physical distancing

After logging in to the test platform, test takers were instructed to open a dummy test containing a few question examples to increase their familiarity and readiness. As previously stated, a pilot test was conducted to gather information for test improvement. Additionally, the test pilot gauged users' experience when using the platform. On average, respondents rated their experience 4.58 out of 5. The favorable results could be attributed to the application's user-friendliness. For test takers from the Ministry of Finance, the application's user interface seems familiar because it uses the same design and theme as the Kemenkeu Learning Center (KLC) online learning platform which is widely used in the ministry. Figures 1 and 2 highlight these platforms' similarities in visual design.



Figure 1 Homepage of Aplikasi Ujian Online



Figure 2 Homepage of Kemenkeu Learning Center (KLC)

Apart from the design similarity, test takers used their Ministry of Finance's single sign-on (SSO) credential when logging in during the scholarship selection. Single sign-on (SSO) is a mechanism that uses a single action of authentication to permit an authorized user to access all related but independent software systems or applications without being prompted to log in again at each of them during a particular session (Radha and Reddy, 2012). This method is convenient as the committee does not have to provide new usernames and passwords since the test takers use the same credentials they use daily to enter the e-office portal. This method, however, is unsuitable for exams of which the test takers are from outside the ministry (e.g., new employees entrance examination). In this case, new usernames and passwords must be generated.

The committee developed two guidelines to ensure that the test was appropriately administered: the general and special circumstances guidelines. The general guideline covered the roles of the central committee and regional committees, test preparations, test delivery, risk mitigation, and infrastructure standards. The special circumstances guideline focused on instructing proctors on the proper course of action if specific problems arose during the test. These problems included non-technical problems caused by the test takers and technical problems due to infrastructure malfunctions.

Proctors/invigilators held a vital role in the tests. Their duties included setting up the computers, checking test takers' identification and PCR/swab test results or vaccination certificates, running the Safe Exam Browser, and monitoring test takers when the tests were running. They were trained through virtual meetings with the central committee. The training also included troubleshooting practices to prepare them to handle potential problems.

The health concern was taken seriously. Before entering the building, the test takers had to pass temperature screening. The proctors would also check the test takers' PCR/swab test results or vaccination certificates. If a participant did not pass the health screening or could not show their PCR/swab test results or vaccination certificates, the proctors would not allow the person to take the test. Once clearing the health screening, the test takers were seated at a minimum distance of 2 meters from others. This arrangement followed the health protocol set by the COVID-19 national task force.

Any dishonesty and noncompliance with the guidelines were also strictly penalized. Two test takers were caught taking notes using scratch paper from the previous Academic Potential Test during the Listening Comprehension section, even after getting a warning from the proctors. It was a violation of the English Competency Test instruction, which explicitly forbids any notes taking. As a consequence, their results were canceled.

Table V Accessibility Challenge

Challenges	Strategies
Test takers with disabilities	Making content colorblind-friendly

Accessibility challenges impact how the test is accessible to a diverse group of test takers. In the past few years, the Ministry of Finance has established a quota for people with disabilities to take the civil service entrance test. Consequently, in designing a test, accessibility for test takers with disabilities should be considered. During the scholarship selection in 2021, there was no record of test takers with disabilities. To an extent, the committee designed the content of the English Competency Test to accommodate certain disabilities. The test, for instance, was designed in black-and-white to accommodate test takers with color blindness because the condition often goes unreported or deliberately be hidden, alternatively termed as 'invisible disability' (Stiles, 2006). Apart from that, the issue of fairness for people with disabilities had not been intensively discussed by the committee and was not addressed in the guidelines. The information regarding disabilities was also not present in the Degree Program Management Technical Policies. This fact mirrored the findings of Lillywhite and Wolbring's study (2019) that demonstrated the lack of information and evidence about disabled individuals that may guide policymakers.

Test Takers' Perceptions

The test takers' perceptions of the test delivery were also investigated using a questionnaire administered through *Aplikasi Ujian Online*. It comprised eight aspects: ease of use, clarity of application user guide, clarity of instruction, clarity of questions, font size, layout, internet stability, and committee responsiveness. The distributions of test takers' perceptions of test delivery are visualized in figure 3.

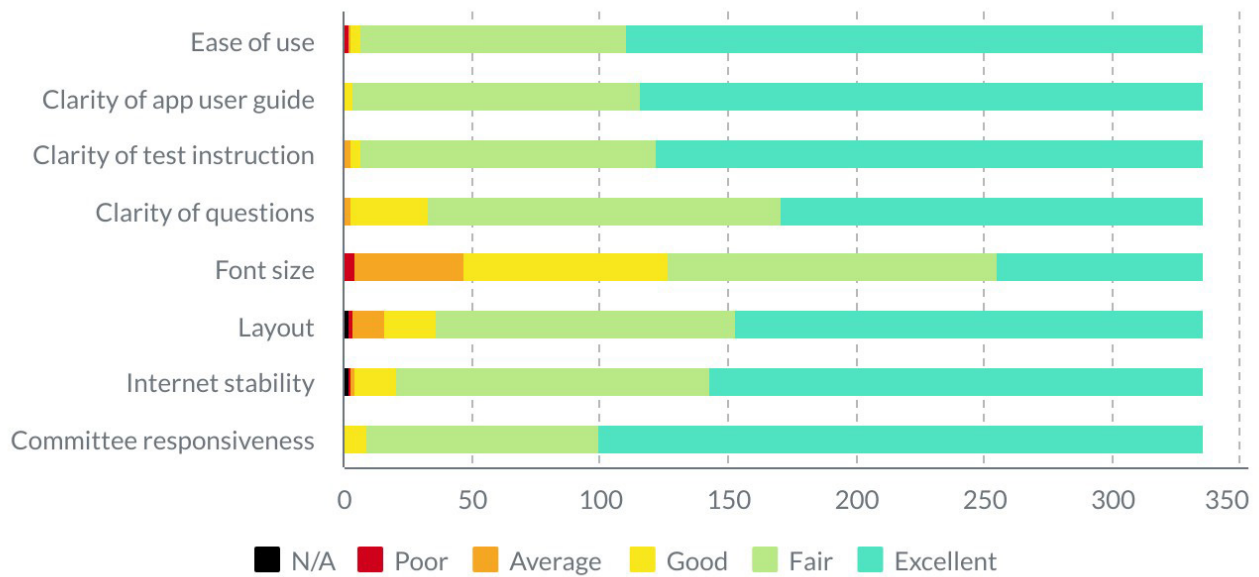


Figure 3 Frequency statistics for questionnaire items on test takers' perception of test delivery

Based on the figure above, the vast majority of test takers expressed their perceptions on all eight aspects as fair or excellent (4 and 5 on Likert scale). There was one missing value indicated with N/A on both ease of use and font size and two missing values both on layout and the internet stability. Given the small number, the missing values accounted for no more than 0.6 percent of the response and were thus insignificant. Figures 4 and 5 show the average response for each aspect.

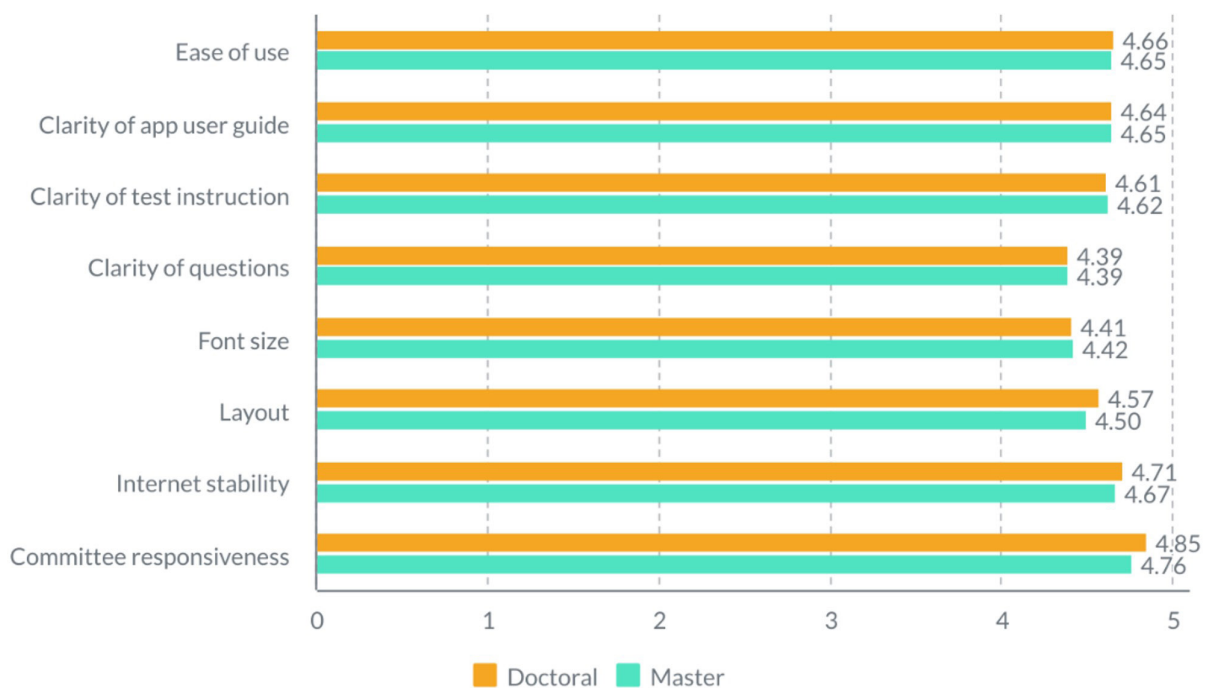


Figure 4 Average of test takers' perceptions by degree

All eight aspects rated 4.60 on average from doctoral participants and 4.58 from master's participants when classified by degree. Two aspects rated less than 4.5 on average. Clarity of questions rated 4.39 from both groups, whereas font size rated 4.41 and 4.42 from doctoral and master's respondents, respectively.

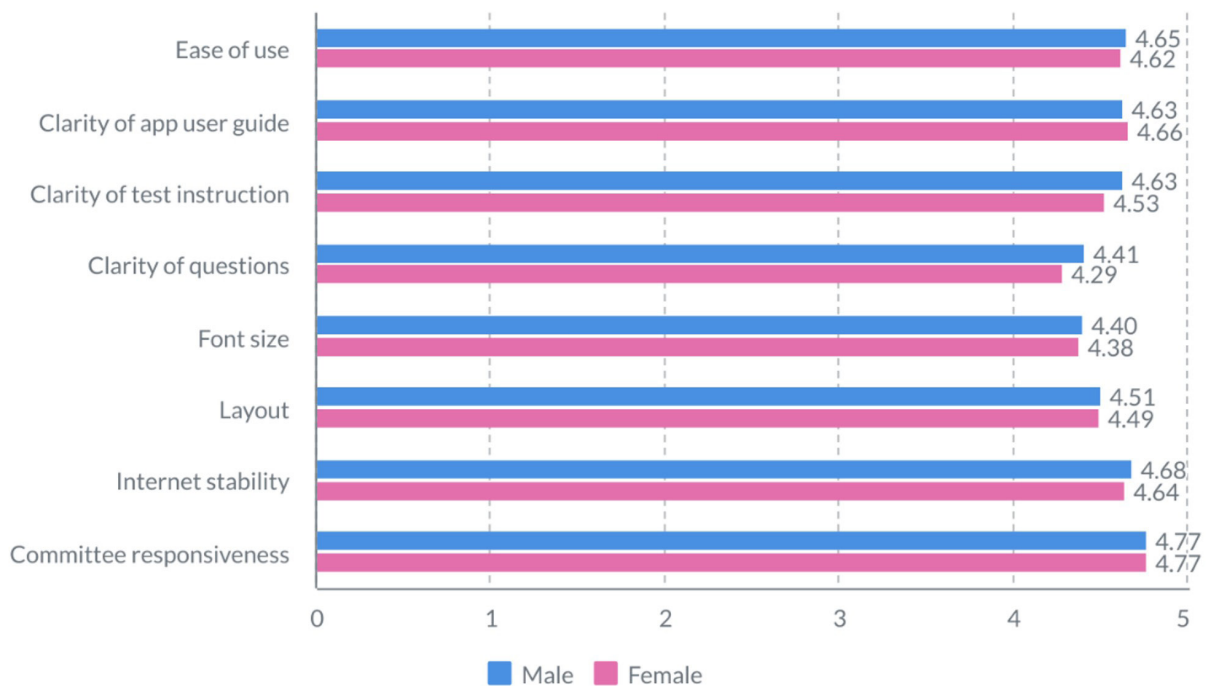


Figure 5 Average of test takers' perceptions by gender

Despite the significant difference in number, female test takers' average response does not differ much from that of male test takers. Male test takers rated an average of 4.59 for all eight aspects, whereas female test takers rated an average of 4.55. This result did not indicate a statistically significant difference from test takers' perceptions by degree. In fact, the outcome visualized in figure 5 virtually paralleled that of figure 4. In summary, test takers from various groups had equivalent perceptions towards the test.

In addition to the quantitative response, we investigated test takers' opinions recorded in the open-ended questions in the questionnaire. Most test takers expressed their satisfaction with the test delivery, noting that the application was easy to use and user-friendly. They learned how to use it swiftly since there was a tutorial video and a practice session with dummy tests before the actual tests. The Ministry of Finance's ID usage as SSO (single sign-on) made it feel seamless since it was the same identification they used daily to access the Ministry of Finance's e-office portal. Test takers also found that the navigation panel was helpful to track the progress of the test, in particular with color-coding to distinguish answered and unanswered questions. The distance between the next and submit buttons was also perceived as far enough to prevent accidental clicks. Overall, they appreciated the CBT delivery in all aspects, particularly the committee whom they found helpful and responsive during the tests.

However, some test takers reported login problems when using 9-digit SSO (older SSO still in use along with the newer 18-digit SSO) and hoped it could be anticipated earlier, especially on the first day of the test. While this problem was unprecedented and therefore it was not covered in the special circumstance guideline, it could be solved swiftly. On the second and third days, the central committee instructed the proctors to ask the test takers before the test whether they still used the older SSO so all test takers could start the test on time.

On the second day, there was a technical glitch that caused the instruction text of the Listening Comprehension section could not be loaded. However, the instruction was also read aloud in the audio.

The problem was solved on the third day of the test. Test takers also suggested that a voice-over be added to the tutorial video since it only used text and background music.

In general, the internet connection in all test centers was perceived to be stable. However, some test takers felt it took too long when loading audio. This problem could be explained since all questions and files had to be downloaded to the client-side before the test started. Once the download was completed, the test could run smoothly.

While most test takers deemed the questions clear enough, some expressed that there were some typos, lack of punctuations, and misnumbered paragraphs. However, they did not elaborate on the examples. Additionally, some test takers suggested that testlets in the Listening and Reading sections could be displayed on one page.

As the aspect with the lowest average value, the font size received mixed responses. Some respondents thought the font size was precisely right, while others thought it was a little too small. One possible explanation for these disparities in perception is that the size of the display has significance. On a wide 22-inch screen, there is an empty space that makes the text appear smaller. Also, older test takers or those with visual problems might perceive the font as small.

Overall, test takers' perception of the computer-based English language assessment was found to be significantly positive. This result is in accord with previous studies that found positive attitudes about other computer-based tests. A possible explanation for this result may be linked to the effectiveness of the strategies to deliver the test reliably and resolve problems rapidly.

Conclusion

Computer-based testing for English competency assessment has both advantages and challenges. Challenges in planning and delivery came from economic, system implementation, test implementation/design, and accessibility aspects. Out of these four, system implementation and test administration challenges were the dominant factors in developing computer-based testing for scholarship selection in the Ministry of Finance. Economically, the implementation of computer-based testing in the Ministry of Finance did not incur substantial additional expenditures and even might save budgets associated with paper-based delivery.

As the preferred mode of delivery, conducting exams at test centers was ideal for high-stake tests of this scale for several reasons. To begin with, the institution's existing computer labs were sufficient to accommodate the test takers. Additionally, the technical issues could be handled better and faster in a controlled environment. Finally, it minimized the incidence of cheating with the combination of human and software proctoring.

To address the challenges, the Ministerial Scholarship committee had proposed a variety of technical and non-technical strategies. Important policies related to the administration of the test were specified in the general guideline and special circumstances guideline to ensure that the test delivery was uniform and that any occurring issues could be resolved. The outcome was reflected in the test takers' perception of the test delivery.

The test takers reported an overall positive perception of the computer-based English Competency Test, indiscriminate of their genders or degrees. These findings were compatible with the findings of recent studies on test takers' perception of CBT. A number of issues during the test were identified, of which the login problem and initial delay were the most frequent. However, these issues could be resolved immediately and had no substantial impact on the test. Responses collected from test takers suggested that a voice-over should be added to the instruction video. Additionally, testlets or groups of related items should be displayed together on the screen.

To ensure fairness for test takers from various backgrounds, policies should be devised to support disadvantaged parties. The absence of information regarding test takers with disabilities suggests that the committee consider creating a guideline for them and provide the assistance they require in the assessment.

Acknowledgement

Acknowledgment is directed to the Human Resource Development Education and Training Center and the Secretariat of Financial Education and Training Agency that supported this research. We are also inevitably thankful for our colleagues at the Division of Integrated Test Management, the Division of Scholarship Management, and the Division of Information Technology and Communication who provided valuable information for this research and made this work possible. Sincere thanks are expressed to Yuniar Widiastuti who became our proofreader and double-checked our work. Finally, a special word of thanks to the late Desi Diarnitha whose kindness and inspiring writings will always remain in our memories.

References

- Alessio, H. M., & Messinger, J. D. (2021). Faculty and Student Perceptions of Academic Integrity in Technology-Assisted Learning and Testing. *Frontiers in Education*, 6, 1–8. <https://doi.org/10.3389/educ.2021.629220>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- CDC. (2018). *Evaluation Briefs No. 16*. <https://www.cdc.gov/healthyyouth/evaluation/pdf/brief16.pdf>
- Chirumamilla, A., Sindre, G., & Nguyen-Duc, A. (2020). Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway. *Assessment and Evaluation in Higher Education*, 45(7), 940–957. <https://doi.org/10.1080/02602938.2020.1719975>
- Frankl, G., Schartner, P., & Jost, D. (2017). The “Secure Exam Environment”: E-Testing with Students’ Own Devices. *IFIP Advances in Information and Communication Technology*, 515, 179–188. https://doi.org/10.1007/978-3-319-74310-3_20
- Khoshshima, H., Morteza, S., Toroujeni, H., & Associate, P. D. (2017). Transitioning to an Alternative Assessment: Computer-Based Testing and Key Factors Related to Testing Mode. *European Journal of English Language Teaching*, 2(1), 54–74. <https://doi.org/10.5281/ZENODO.268576>
- Lillywhite, A., & Wolbring, G. (2019). Undergraduate Disabled Students as Knowledge Producers including Researchers: A Missed Topic in Academic Literature. *Education Sciences* 2019, 9(4), 259. <https://doi.org/10.3390/EDUCSCI9040259>
- Medina, M. S., & Castleberry, A. N. (2016). Proctoring strategies for computer-based and paperbased tests. *American Journal of Health-System Pharmacy*, 73(5), 274–277. <https://doi.org/10.2146/ajhp150678>
- Nassaji, H. (2015). Qualitative and descriptive research: Data type versus data analysis. *Language Teaching Research*, 19(2), 129–132. <https://doi.org/10.1177/1362168815572747>
- Radha, V., & Reddy, D. H. (2012). A Survey on Single Sign-On Techniques. *Procedia Technology*, 4, 134–139. <https://doi.org/10.1016/J.PROTCY.2012.05.019>
- Stiles, J. (2016). Colorblindness: Invisible Disability. *Iowa Science Teachers Journal*, 33(1), 19–22.
- Sulistiyono, R. S., Suyata, P., & Rahayu, T. (2018). Inventory Survei Ujian Nasional Bahasa Indonesia Berbasis Komputer. *Deiksis - Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 5(1), 69–80. <https://doi.org/10.33603/deiksis.v5i1.920>
- Sulastri, S., & Puspawati, I. (2019). A Computer-Based Standardized Testing: Its Challenges and Strategies. *Journal of Foreign Language Teaching and Learning*, 4(1), 14–27. <https://doi.org/10.18196/ftl.4135>
- Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). *Computer-based Testing: Practices and Considerations*. National Center on Educational Outcomes, University of Minnesota. <https://nceo.umn.edu/docs/onlinepubs/synthesis78/synthesis78.pdf>
- Winke, P. M., & Isbell, D. R. (2017). Computer-Assisted Language Assessment. In *Language, Education and Technology* (pp. 313–325). Springer International Publishing. https://doi.org/10.1007/978-3-319-02237-6_25