Understanding the implementation of an at-home language test: A case of an online version of TOEFL-PBT

Refanja Rahmatillah[*1], Rizki Fajrita[2], Endah Anisa Rahma[1]

[1]Universitas Teuku Umar, Indonesia
[2]Universitas Syiah Kuala, Indonesia

ABSTRACT

An at-home test is a unique mode of language test delivery as a result of mass-gathering prohibition during the COVID-19 pandemic. Despite the uniqueness, little is known about how to effectively implement an at-home test. This study aims to provide a deeper understanding of the test by exploring the execution of the online version of TOEFL-PBT in the Language Center of Syiah Kuala University. Four test administrators were interviewed to share their experiences and opinions related to considerations for implementing an at-home proficiency test, which includes technological resources, security, and validity concerns. The data were then analyzed descriptively. The results of this study revealed that the Language Center used Safe Exam Browser to deliver the test and Zoom to supervise the test-takers in real time. The proctors could stop the test and privately investigate the test takers using the Zoom Breakout feature. The validity of the test was claimed not to be a concern since the test provider used the same form of questions as the offline version. In addition, the Language Center expressed exhaustion in carrying out the online test, thus suggesting the development of a less complicated procedure of an at-home test.

Keywords: *At-home test; TOEFL; Technological resources; Security; Validity*

---

[*]*Corresponding Author:*

Refanja Rahmatillah
Universitas Teuku Umar
Jl. Alue Peunyareng, Gunong Kleng, Kec. Meureubo, Kabupaten Aceh Barat, Aceh 23681, Indonesia
Email: refanjarahmatillah@utu.ac.id

1. Introduction

The novel coronavirus or COVID-19 pandemic has changed the education system worldwide from a face-to-face to an online environment (Gacs et al., 2020), including in the implementation of language testing (Isbell & Kremmel, 2020). An at-home language test has become a new instrument to measure someone's language ability in this restricted situation and this phenomenon requires test providers and users to adjust the execution of the test (Plough & Raquel, 2020). These changes, unfortunately, potentially create a high risk of abandonment of the test principles, format, and content, which may result in less validity and reliability of the test (Hrbackova et al., 2020).

Before COVID-19 outbreaks, the concern related to different formats of test delivery, paper-based testing (PBT) versus computer- or internet-based testing (CBT or iBT), has long arisen. Similar to the concern mentioned above related to the changes of the testing procedure to be at-home supervised, the shift from PBT mode to CBT/iBT ones also results in the studies exploring the issues of test-taker's performance, reliability, and validity of the test, and test-taker's perceptions related to these two modes of test administration (Brunfaut et al., 2018; Hosseini et al., 2014; Öz & Özturan, 2018). Brunfaut et al. (2018), for example, conclude that the different modes of the test (PBT and CBT) impact the performance of the writing test-takers, even with a small effect. Öz and Özturan (2018), on the other hand, report that both versions are highly reliable and valid and have no significant difference in the test score achieved by the participants. In addition, Hosseini et al. (2014) achieve a not conclusive finding that slightly more than half of the test-takers prefer CBT because of the enjoyable and time-saving features, while the others claim that they are not familiar with CBT and that the use of computers tires their eyes. Thus, based on these studies, it can be assumed that there is no agreement on whether different modes of test can affect the overall execution of the test.

The previous studies mentioned, however, may not comprehensively answer the current issues of test implementation during the COVID-19 pandemic. It is because the past two types of tests, PBT and CBT, were both carried out with in-person supervision. On the contrary, due to the prohibition of mass gathering, the tests conducted at the current time have to be online supervised. Despite this unique condition, only a few studies had examined the execution process of at-home tests. A study by Isbell and Kremmel (2020) reviews several language testing centers that implement at-home tests and then emphasizes the importance of considering technology resources, security, and validity to carry out such tests. These considerations are crucial for guaranteeing that at-home tests can be professionally carried out as in-person supervised tests.

By referring to the crucial points delivered by Isbell and Kremmel (2020) related to considerations in implementing an at-home language test, this study tries to explore the actualization of the considerations in an at-home TOEFL test. This study tries to answer the following questions:

1. How does a language testing center carry out an at-home test?

2. Does the language center consider technology, security, and validity concerns in implementing an at-home test?
3. What are the challenges faced by the language testing center in executing the at-home test?

2. Literature review

2.1. Technology in language testing

The incorporation of technology in language testing, started with the introduction of CBT, is deemed to be able to provide some advantages, including easier test distribution and faster and more accurate scoring (Scheuermann & Björnsson, 2009). Nevertheless, one obvious problem with the implementation of CBT is the concern with computer familiarity. Even though studies like Taylor et al. (1999) and Khoshsima et al. (2019) reveal that test-takers who are low- and high-computer-familiar produce no meaningful differences in the test scores and Dooey (2008) claims that the issue of computer familiarity seems less prevalent at this time, test-takers may still show anxiety that their computer familiarity might affect their performance (Yu, 2010).

Even when iBT emerged where it is perceived as more convenient and safer to deliver items compared to CBT (Roever, 2001), iBT still contributes to several substantial challenges encountered by the test takers. Barkaoui (2015) reports constraints related to keyboarding and the typographical error committed during the TOEFL iBT examination, especially in the writing section, and Wolfe and Manalo (2004) show that some participants who express concern related to computer familiarity tend to choose handwriting essays over word processing. They perceive the utilization of a computer might trigger nervousness while typing and worsens their performance. However, there is still a debate on whether the computer familiarity concern brings about a negative influence on iBT test-takers as according to Weigle (2010), more regions and communities are exposed to the digital-centered environment, and more TOEFL iBT by ETS placements are delivered worldwide.

Looking deeper into the utilization of technology, the aspect of practicality is inevitable to be taken into account. Though CBT provides a better environmental impact compared to paper-based tests, it is still far from perfection, and the procurement and maintenance cost of the infrastructure for long-term viable usage is quite pricey (Nogami & Hayashi, 2010). The shift from CBT to iBT in language testing is believed to tackle the said issue since the iBT programs utilize a platform instead of hardware that is prone to wear and tear (Laborda et al., 2010). Some technical problems, however, might obstruct the progress of implementing a web-based test. Roever (2001) points out some glitches probably occur during a running online test that includes buffering loading time, frozen pages caused by server traffic, and depending on the test-taker's computer speed, such a problem would even worsen. iBT or a web test, moreover, would sometimes turn from inexpensive resolution for administrators to a costly choice for test users as access to the internet and computer device is considered luxurious for some (Isbell & Kremmel, 2020).

2.2. Security in language testing

A proper testing practice should be also capable of ensuring reliable security for preventing possible breaches in its system that can unfavorably impact the examination credibility. When technology is embedded in a language testing, its security somehow gets more vulnerable to breaches, and concerns on how to enable maximum and reliable security system for preventing possible threats, especially in terms of the leaks of test items to the public, rise significantly (Ockey, 2009; van der Linden & Glas, 2000). An instance of a high-stake test leak case happened in 2002 when some Asian language websites in which the questions along with the answers of computer-based Graduate Record Examination (GRE) conducted by ETS were shared with its user (Wired, 2002). This circumstance happened due to the limited variety of item banks, and even when ETS is renowned for its tremendous resources, it still once failed to maintain its security.

Another challenge in incorporating technology in tests lies in tackling identity fraud, especially in iBT, where it is possible to complete the test independently without attending the test center. Ockey (2009) argued this system may lead to a condition in which the one who registered for the system might susceptible employ a qualified person to sit on the test if the security of the test program is compromised. To prevent such cheating attempts, the administrator can apply a secure mode or a third application on every test to monitor the screening time to confirm the participant's identity. This method can also lessen the probability of cheating by completing assigned tasks such as browsing for answers and asking for an assist from other individuals (Roever, 2001).

2.3. Validity in language testing

The incorporation of technology in language tests also raises the concern of its validity as Suvorov and Hegelheimer (2014) claim that after the changes took place, construct validity is the aspect that is mostly researched. Brown and Abeywickrama (2010) define construct validity in language testing as the relevance of the test in measuring language skills as the language ability theories expect. Construct validity is crucial in the execution of CBT/iBT in guaranteeing that the test measures language skills, not computer skills (Dooey, 2008).

Most of the studies that investigate the validity issues in CBT/iBT carry out the validation by comparing the CBT/iBT scores and PBT scores (Boonsathorn & Kaoropthai, 2016; Bunderson et al., 1988; Coniam, 2006; Piaw, 2012). Bunderson et al. (1988), as the earliest work, and Coniam (2006), both find that test takers achieve better but not significant results in the computerized test compared to paper administration. This finding indicates that both paper- and the computer-based test have similar construct validity. Piaw (2012) in his study which particularly investigates the validity issues in CBT conclude that CBT possesses high validity in terms of test performance and can be used as a substitute for the PBT version. Meanwhile, the exploration of validity in iBT is carried out by Boonsathorn and Kaoropthai (2016), which examines the web-based mC-

test. They conclude that the web-based test has high validity, both criterion-related and face validity, and presume that web-based tests can be an alternative to the traditional mode of testing.

Despite the consistency of the findings that reveal that CBT/iBT is as valid as PBT and, thus, can be used as an alternative, one consideration should be noted that the promotion of performance-based assessment in language learning cannot be presumably easily done via computer administration. Chalhoub-Deville (2001) highlights this concern and argues that performance-based assessment is conducted better via paper-based administration. Even though the emergence of iBT tries to solve this problem, the use of technology to administer performance-based assessments will still generate problems in terms of practicality (Chalhoub-Deville, 2012).

2.4. At-home language test

While there are numerous studies that answer the concerns for the implementation of CBT and iBT, they may not fully apply for the new method of test, at-home test, since it is not in-person supervised. An at-home language test is delivered online and can be completed at the test-taker's own house (Isbell & Kremmel, 2020). While it is the same as iBT in terms of the requirement of fast and stable internet connection, the use of headphones may be tricky for at-home tests. When the use of headphones is supposed to hinder the test takers to record or letting others listen to the audio material, at-home test takers test can contrarily use headphones to receive help from others Wagner (2020). This problem can actually be solved by applying video-based proctoring where test takers must show themselves and their room throughout the test or using exam security software which may temporarily restrict the functionality of the test-taker's computer and allows proctors to monitor the activity on the computer (Isbell & Kremmel, 2020).

Due to their prone to cheating, validity becomes one of the highlighted issues in an at-home test. Unfortunately, only two studies to date have compared the scores from at-home tests and other test modes to examine the validity concern, Rigo (2020) and Stradiotová et al. (2021). These two studies share the same findings where test-takers with paper-based mode achieve better scores than those taking the at-home test. These findings, however, are not enough to validate that an at-home test has lower validity compared to other test modes, and more studies need to be conducted to validate this finding).

3. Method

3.1. Participants and location

This study employed a qualitative method with an interview technique to answer the research questions. Four staff at the Language Center of Syiah Kuala University, two males and two females, who are involved in the delivery of an at-home TOEFL test were interviewed. The Language Center was chosen because it is, as far as the researchers' concern, the only testing center in Aceh, Indonesia that carries out an at-home test. The

four staff were recruited since they were all that deliver the at-home TOEFL test. They began to deliver online remote proctoring for both TOEFL ITP and TOEFL-Equivalent tests in July 2020. Previously, the tests were delivered on-site in paper-based mode.

3.2. Data collection

The four test administrators were interviewed virtually via Zoom. The focus group interview was applied to obtain affirmed information among the participants (Short, 2006), and a semi-structured interview was used to provide flexibility for the participants in giving answers, thus resulting in richer data (Richards, 2009). The interview questions were formulated based on Isbell and Kremmel's (2020) considerations in administering an at-home test. There were 17 questions in total, consisting of five questions asking about a technological concern, eight questions about security concern, two questions about validity concern, and the other two questions about challenges that the center faced during the implementation of the at-home test. The interview was carried out in Indonesian language and lasted for approximately 40 minutes.

3.3. Data analysis

The interview was recorded to be then analyzed. The analysis followed Richards' (2009) guidelines in analyzing and interpreting interview data. The analysis began by transcribing the interview, followed by finding phrases that respond to the questions, and ended by categorizing the findings under the same themes.

The analysis started with the exploration of the participants' answers related to the general execution of the at-home TOEFL test. Then, the answers covering the technological, security, and validity concerns were grouped into respective themes. Finally, the answers expressing the center's struggle during the at-home implementation, as well as their expectation for future application, were gathered in one theme. The complete explanations of the findings are presented below.

4. Findings

This part explores the results of the interview with the four test administrators. As mentioned in the method part, there were 17 questions asked, and the responses to these questions provided the answer for the three research problems addressed in this study.

4.1. The execution of at-home test

This part answers the first question of this study which is about the administration of an at-home test. In general, the Language Center of Syiah Kuala University carried out the at-home TOEFL through several processes. The first process was the process to move the paper-based test to an online administration. Here the administrators conveyed that they chose Moodle to design the test and Safe Exam Browser (SEB) to deliver it. "We used Moodle to design the questions of the test because it was the simplest application to use. When we finished the design, we put them on the Safe Exam Browser" (Test Administrator 1).

Once the test set was ready to be delivered, the test was open for registration and the test-takers were required to do a mini-test simulation provided by the administrators. This simulation process proposed to check the qualification of the test-takers to take the test and to give insights to the test-takers of what the test would be like. "There was a test simulation held one day before the actual test to check the test-takers' internet speed. When they did not meet the   Internet speed standard, they could not take the online test" (Test Administrator 2).

Those who were qualified to take the online test would be given a password to be entered on SEB via email. On the next day, before taking the test, the identity of test-takers would be firstly checked via Zoom by asking them to face the camera while the administrators matched them with the data registered. The test-takers were also required to show their room to prevent the presence of illegal materials or help. After that, the test-takers would complete the test on their computer while being supervised via Zoom opened from their phones. "The test-takers were required to open Zoom via their phones and put the phones behind them, in a position that we could see their whole rooms" (Test Administrator 3).

When the test proctors presumed something suspicious happened, they could stop the test on SEB and put the test-taker into a Zoom breakout room to be investigated. "We, proctors, could see something suspicious happening via Zoom. If that happened, we stopped the test and used the Zoom breakout room feature to investigate the test-taker" (Test Administrator 2).

Finally, when the test-takers completed the test, they would later receive their scores via e-mail.

4.2. Considerations in implementing at-home test

The explanation of the administration processes of the at-home TOEFL above, actually, describes how the testing center considered the technology and security concerns in implementing the test but paid little attention to validity issues, which answers the second research problem of this study. The detailed description is presented below under each concern.

4.2.1. Technological concern

The use of SEB to deliver the test was based on some reasons. Besides a well-known platform to securely administer a test, the administrators stated that SEB can be well run in almost all operating systems, except for Windows 8. "SEB can be well accessed via almost all operating systems, except for Windows 8 that it would run slowly" (Test Administrator 1).

In addition, to be able to take the at-home test, the administrator required the test-takers to have a personal computer to run SEB, a phone to open Zoom, earphones, and at least 3GB of fast and stable internet connection. To ensure the satisfaction of this requirement, the administrators held a mini-test simulation on the day before the test to check the test takers' internet speed. "It is a must for the test-takers to have a fast and stable

internet connection. That's why we conducted an internet bandwidth test one day before the test date" (Test Administrator 3).

Despite the stated preparation to handle the technology concerns, problems did occur especially with the use of a phone for Zoom supervision. "The problem that mostly occurred on the test day was the test-takers' phones became overheated and this disturbed the supervision process, thus we needed to stop the test" (Test Administrator 4).

4.2.2. Security concern

As described in the processes of test administration earlier, the testing center considered the security issues through some steps. To prevent the occurrence of identity fraud, before the test, the identity and the room of test-takers would be checked by using Zoom:

> Before starting the test, we checked the test-takers' identity. We asked them to close their faces to their phone camera and we would match their faces with the ones registered. After that, we also asked them to show their whole rooms by rotating their phones (Test Administrator 2).

In relation to preventing test item breaches, the administrators believed that the use of SEB and the requirement to use earphones could minimize the risk:

> The features in test-takers' computers could not be opened during the use of SEB thus they could not export the test items. Moreover, we required them to use earphones during the listening section so they could not record the audio nor let others listen to it (Test Administrator 1).

In addition, the use of Zoom to supervise the test-takers in real-time prevented them from cheating. "By using Zoom we could monitor the activities of the test-takers. We also knew if there was somebody else entering their room to help them" (Test Administrator 4).

Finally, the examiners could stop the test if they found something suspicious and used the Zoom Breakout feature to investigate the test takers.

4.2.3. Validity concern

The validity issue seemed to be the one that was less noticed by the test administrators. There was no comparison carried out by the testing center between the scores obtained from the at-home and traditional paper-based TOEFL since they believed that the transformation of the valid PBT TOEFL to online administration would not change the validity quality. "We know that we use valid TOEFL questions, but since the administration of at-home TOEFL we have not compared the test-takers' scores from the two modes of delivery" (Test Administrator 1).

Moreover, the test results were also issued in the same format as the traditional mode of delivery where the administrators did not mention that the scores were obtained

through an at-home test. "We did not differentiate the results between at-home and paper-based TOEFL. We issued the results in the usual format" (Test Administrator 4).

4.3. Challenges in implementing at-home test

Regarding the last research problem addressed in this study, the administrators revealed that they faced some challenges in implementing the at-home TOEFL. The first challenge that they stated was that the amount of effort in carrying out the at-home test was not worth the number of test-takers participating in the test and, as a result, the income received:

> We do not think the income we received, which was the same as administering the PBT TOEFL, was worth the effort that we spent. It is because the number of test-takers that can be accommodated was fewer than the usual test (Test Administrator 1).

Another challenge conveyed by the administrators was that they had to use their personal devices and sometimes their own internet data. This burdened the staff and they hoped for the procurement of devices for the staff:

> I personally hope that the office can provide devices for the staff so that they don't need to use their own devices. Moreover, when we worked from home, we also had to use our own internet data (Test Administrator 3).

5. Discussion

The results of the interview with four at-home TOEFL administrators of the Language Center of Syiah Kuala University provide insights to answer the research questions addressed in this study. The testing center carried out the test by using Safe Exam Browser (SEB) and Zoom and held a mini simulation test one day before the real test. They did consider the technology and security demands, but paid less attention to validity concerns, and faced some challenges in carrying out the at-home TOEFL. From the point of technology demands, as pointed out by Isbell and Kremmel (2020), a fast and stable internet connection is a crucial feature in the implementation of an at-home test. The testing center provided a quite fast and stable internet connection for the administrators to run the test and this rarely became a problem. In addition, the simulation held one day before the real test not only proposed to check the test takers' internet speed and stability but also to familiarize the test takers with the platform used for the test. While the test administrators' narration affirms the findings from previous studies (Dooey, 2008; Hosseini et al., 2014; Khoshsima et al., 2019) that computer familiarity is no longer a problem in this time, such technical problems did occur a few times, e.g. test-takers did not know how to download the test platform app and fill in the required passwords. Thus, the simulation provided assistance for those who were not familiar with the platform and might reduce their anxiety (Yu, 2010) in completing the real test.

Furthermore, the accessibility of the technology used in the delivery of the test seems to be handled well. The use of SEB to deliver the test is said by the administrators to run well in almost all computer operating systems. The utilization of such a platform that can be fairly assessed will likely result in equal opportunities for the test-takers in demonstrating their language proficiency (Laborda et al., 2010). Even though some studies, such as (Roever, 2001) and (Stradiotová et al., 2021), raise a concern related to the accessibility of an online test delivery due to limited internet coverage in some parts of the world, this issue will be likely solved in the future for more and more regions start to have internet access.

On the other hand, practicality appeared to become an issue in the implementation of the at-home TOEFL. Since the administrators required the test takers to use phones to open the Zoom app for the live supervision purpose, it is reported that the phones frequently became overheated and then automatically turned off, and, as a result, it disturbed the test progress. It is when Dooey (2008) reminds us that the equipment used in a test should be in perfect working condition to prevent such problems from occurring. Moreover, aligned with the concerns conveyed in Nogami and Hayashi (2010) and Ockey (2009) that costs are the issue in conducting a test with technology, the administrators admitted that they had to expend more money and effort in delivering the at-home TOEFL. The amount of income from the test fee, however, was viewed as not sufficient to cover the expense since the number of test-takers that could be accommodated was fewer than in the normal test condition. This might be an interesting finding for it contradicts the argument that the emergence of iBT can accommodate a greater number of test-takers at one time (Roever, 2001).

Moving to the point of security concerns, the test administrators appeared to well address the security issues raised in the incorporation of technology in language testing. By applying live video proctoring using Zoom, the proctors could monitor the examinee's activities and environment during the test, which is a solution to overcome the shortcoming of direct monitoring in an online test (Isbell & Kremmel, 2020). As a result, the worry of cheating that is prone to happen in at-home tests (Wagner, 2020) could be also minimized. In addition, the video conference app was also used to prevent the possibility of identity fraud conducted by the examinees. As pointed out by Ockey (2009), identity fraud commonly occurs in iBT since the test takers do not need to come to the test center to take the test. The use of such an application to screen the test takers before starting the test, thus, will overcome this issue.

Regarding the concern of test item breaching, as the main security concern in the administration of CBT or iBT (Ockey, 2009; van der Linden & Glas, 2000), the test administrators believed that the use of SEB as the test platform and the requirement of using earphones during listening section were able to minimize the risk. SEB could stop the functions of other applications on the examinees' computers; thus, they would not be able to duplicate or screenshot the test page. Meanwhile, the use of headphones would prevent the examinees from recording or let other people listen to the test audio.

Unfortunately, the validity concern appeared to be less evaluated by the administrators in implementing the at-home TOEFL. This happened due to two factors, the belief in the equally valid quality of the test since it used the same form of questions as the paper-based version and the absence of test takers' scores both in paper-based and at-home TOEFL for comparison. Nevertheless, to guarantee that the test measures the intended measurement validity checking is a must in CBT/iBT (Dooey, 2008) and most previous studies accomplished it by comparing the scores of two different modes of the test (Suvorov & Hegelheimer, 2014). Therefore, it is an urgent need to measure the validity of the scores considering that the test center also did not explicitly mention in the score reports that they were obtained through an at-home mode, which is essential information for score users (Isbell & Kremmel, 2020).

Further challenges were faced by the test administrators, besides those mentioned earlier related to the practicality issues. The test administrators revealed that administering an at-home test is more complicated than the traditional one in terms of preparing the test and communicating the test procedures to the test takers. It is when Ockey (2009) points out the necessity of human resources who are competent in using technology. Moreover, while proctoring the at-home TOEFL can be done at home, the consequences of using personal devices and internet data burden the test proctors. Hence, less complicated procedures and the availability of equipment for test proctors are needed to take into account in implementing the effective at-home test.

6. Conclusion

This study aims to provide an understanding of the implementation of an at-home test, as a method to deliver language tests during this COVID-19 time, by reviewing the case of an at-home TOEFL carried out by the Language Center of Syiah Kuala University. In short, the testing center transformed the PBT version of the test to be delivered digitally using SEB and used Zoom to remotely supervise the test-takers. In implementing these processes, the testing center appeared to consider both technology and security concerns but failed to guarantee the validity of the test. In addition, they also faced challenges in carrying out the at-home test, such as the imbalance between the effort spent and the income received and the use of personal devices. They hope for a less complicated procedure to carry out the at-home test, which further research may accomplish.

Furthermore, this study serves as a starting point to explore the other aspects involved in an at-home test. For example, further research can study the effect of the use of such remote technology on test-takers' performance or investigate the validity of the at-home test by comparing test-takers' scores with other test modes. These explorations might be crucially needed in consideration that future language testing administration still has to be online proctored because of the unresolved pandemic situation.

References

Barkaoui, K. (2015). Test takers' writing activities during the TOEFL iBT® writing tasks: A stimulated recall study. *ETS Research Report Series*, *2015*(1), 1–42. https://doi.org/10.1002/ets2.12050

Boonsathorn, S., & Kaoropthai, C. (2016). QSAT: The web-based mC-test as an alternative English proficiency test. *TESOL International Journal*, *11*(2), 91–107.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices.* Pearson Education.

Brunfaut, T., Harding, L., & Batty, A. O. (2018). Going online: The effect of mode of delivery on performances and perceptions on an English L2 writing test suite. *Assessing Writing*, *36*(September 2017), 3–18. https://doi.org/10.1016/j.asw.2018.02.003

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1988). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., Issue June 1988, pp. 367–407). Macmillan.

Chalhoub-Deville, M. (2001). Language testing and technology: Past and future. *Language Learning and Technology*, *5*(2), 95–98.

Chalhoub-Deville, M. (2012). Technology in standardized language assessments. In R. B. Kaplan (Ed.), *The oxford handbook of applied linguistics, (2 Ed.)* (2nd ed., Issue September 2018, pp. 1–19). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195384253.013.0035

Coniam, D. (2006). Evaluating computer-based and paper-based versions of an English-language listening test. *ReCALL*, *18*(2), 193–211. https://doi.org/10.1017/S0958344006000425

Dooey, P. (2008). Language testing and technology: Problems of transition to a new era. *ReCALL*, *20*(1), 21–34. https://doi.org/10.1017/S0958344008000311

Gacs, A., Goertler, S., & Spasova, S. (2020). Planned online education versus crisis-prompted online language teaching: Lessons for the future. *Foreign Language Annals*, *53*(2), 380–392. http://clik.dva.gov.au/rehabilitation-library/1-introduction-rehabilitation%0Ahttp://www.scirp.org/journal/doi.aspx?DOI=10.4236/as.2017.81005%0Ahttp://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/as.2012.34066%0Ahttp://dx.doi.org/10.1016/j.pbi.2013.02.0

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

Hrbackova, J., Lukacova, E., & Puncocharova, E. (2020). Language testing in the time of the Covid-19 pandemic 2020. *CASALC Review*, *10*(1), 186–189.

Isbell, D. R., & Kremmel, B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, *37*(4), 600–619. https://doi.org/10.1177/0265532220943483

Khoshsima, H., Toroujeni, S. M. H., Thompson, N., & Ebrahimi, M. R. (2019). Computer-based (CBT) vs. paper-based (PBT) testing: Mode effect, relationship between computer familiarity, attitudes, aversion and mode preference with CBT test scores in an asian private EFL context. *Teaching English with Technology*,

19(1), 86–101.

Laborda, J. G., Magal-Royo, T., & Carrasco, E. E. (2010). Teachers' trialing procedures for computer assisted language testing implementation. *Eurasian Journal of Educational Research*, *39*, 161–174.

Nogami, Y., & Hayashi, N. (2010). A Japanese adaptive test of English as a foreign language: Developmental and operational aspects. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 191–211). Springer. https://doi.org/10.1007/978-0-387-85461-8

Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *Modern Language Journal*, *93*(SUPPL. 1), 836–847. https://doi.org/10.1111/j.1540-4781.2009.00976.x

Öz, H., & Özturan, T. (2018). Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies*, *14*(1), 67–85. https://dergipark.org.tr/en/download/article-file/650156

Piaw, C. Y. (2012). Replacing Paper-based testing with computer-based testing in assessment: Are we doing wrong? *Procedia - Social and Behavioral Sciences*, *64*, 655–664. https://doi.org/10.1016/j.sbspro.2012.11.077

Plough, I., & Raquel, M. (2020). *ILTA Newsletter, 4*(1). https://www.iltaonline.com/page/NewsletterMay2020

Richards, K. (2009). Interviews. In H. Juanita & C. Robert A. (Eds.), *Qualitative research in applied linguistics* (pp. 182–199). Palgrave Macmillan.

Rigo, F. (2020). Effectiveness of testing English online in distance. *Marketing Identity*, *1*, 547–557.

Roever, C. (2001). Concerns with computerized adaptive oral proficiency assessment. *Language Learning & Technology*, *5*(2), 84–94. http://llt.msu.edu/vol5num2/roever/default.html

Scheuermann, F., & Björnsson, J. (2009, November 25). *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities . https://doi.org/10.2788/60083

Short, S. E. (2006). Focus group interview. In P. Ellen & C. Sara R. (Eds.), *A Handbook for social science field research: Essays & bibliographic sources on research design and methods* (pp. 103–116). Sage Publications.

Stradiotová, E., Nemethova, I., & Stefancik, R. (2021). Comparison of on-site testing with online testing during the covid-19 pandemic. *Advanced Education*, *8*(18), 73–83. https://doi.org/10.20535/2410-8286.229264

Suvorov, R., & Hegelheimer, V. (2014). Computer-assisted language testing. In A. J. Kunnan (Ed.), *The companion to language assessment* (first, pp. 594–613). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118411360.wbcla083

Taylor, C., Kirsch, I., Eignor, D., & Jamieson, J. (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

van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized adaptive testing: Theory*

and practice. Kluwer Acadmic Publishers.

Wagner, E. (2020). Duolingo English test. *Language Assessment Quarterly*, *17*(3), 300–315. https://doi.org/10.1080/15434303.2020.1771343

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing*, *27*(3), 335–353. https://doi.org/10.1177/0265532210364406

Wolfe, E. W., & Manalo, J. R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning and Technology*, *8*(1), 53–65.

Yu, G. (2010). Effects of presentation mode and computer familiarity on summarization of extended texts. *Language Assessment Quarterly*, *7*(2), 119–136. https://doi.org/10.1080/15434300903452355