Thai Learners' Performance on Listening Test: A Comparison of Paper-based and Web-based Testing*

Sujane Panjan**
Thanyapa Palanukulwong***

Abstract

Web-based testing has been used in many educational institutions, but it is argued that the performance of test takers is related to the test delivery mode and their preferences. This study aimed to investigate whether the use of traditional paper-based tests and web-based tests to test listening ability would affect test takers' performance. Two parallel tests were administered to 200 Thai secondary students on two different occasions. The test takers were first administered the paper-based listening test. Then, the web-based listening test was administered one week after the first test. Paired sample t-test was used to compare the means of two test modes. The result of the study showed that the difference in test delivery mode did not affect overall participants' listening test performance. However, there were significant differences in the test scores between paper-based and web-based testing modes between the high and low proficiency groups. The high group had a better score on the paper-based listening test than the web-based test; the low group achieved significantly higher scores on the web-based listening test than on paper-based. The study also revealed that test mode preferences might not reflect their test performance. Guidelines and suggestions for test developers and teachers that may help in developing a listening test for students with different characteristics, as well as preferences in learning and testing styles were provided.

Keywords: Listening Test, Paper-based Testing, Web-based Testing, Test Performance, Test Preference

* This article is conducted to compare the students' performance on listening test between paper-based and web-based testing and to fulfill the requirement for Master of Arts in Teaching English as an International Language, Faculty of Liberal Arts, Prince of Songkla University, Hatyai Campus.

** M.A. student, Faculty of Liberal Arts, Prince of Songkla University, Hatyai Campus, E-mail: sujane@spt.ac.th

*** Assoc. Prof. Dr., Faculty of Liberal Arts, Prince of Songkla University, Hatyai Campus, E-mail: thanyapa.c@psu.ac.th

Veridian E-Journal, Silpakorn University
ISSN 1906 – 3431

Introduction

Listening is one of the most important fundamental language skills for human beings. People spend 40-50% of their communication time on listening and spend the rest on speaking, reading, and writing (25-30%, 11-16%, and 9%) respectively (Mendelsohn, 1994). Listening is also a central to the lives of students throughout all levels of educational development because it plays an important role in facilitating of learning (Coakley & Wolvin, 1997; Wu, 2005). Students from kindergarten through high school are expected to listen 65-90 percent of their learning time and both in and out of the classroom, listening consumes more of daily communication time than other form of verbal communication (Gilbert, 1988; Wolvin and Coakley, 1988). Without listening skills, learners never learn to communicate effectively (Nunan, 1998).

However, Listening is a highly complex mental process that involves perception, attention, cognition, and memory. Listeners have to employ all types of knowledge to interpret meaning, identify the level of stress and intonation, understand vocabulary and grammatical structures, recognize and remember what they have heard (Byrnes, 1984; Vandergrift, 1999). Therefore, in language acquisition, comprehending sound and speech in a foreign language is quite a difficult task for EFL learners. Due to the differences between the phonological system and linguistic structure of L1 and L2, foreign language learners must process the incoming information using linguistic rules and patterns that they have not mastered (Shang, 2008; Thomas & Dyer, 2007).

To measure listening comprehension, most teachers use listening test to evaluate the listening proficiency. However, listening is not an easy skill to test because it is less convenient and more technically complicated than conducting reading or writing tests. Listening tests require suitable input material with a good quality of sound system which can affect understanding of listening (Conaim, 2006). The listening room should support several acoustic functions and the sound system should transmit voices to the listeners via both direct and reflected sounds. According to Waye et al. (2015), the direct sounds are most important for listeners who sit close to the speaker and reflected sounds are more important for distance listeners. However, in generally, most classrooms do not have multimedia systems and suitable resources such as computer, smart board and listening laboratory to support listening tests.

With the rapid development of Information and Technology (IT), computer-based tests have been used in language testing. Since early 80th, computers as a testing medium attract the attention of psychometricians because they allow the application of items response

theory for delivering adaptive tests, which can often pinpoint a test taker's ability level faster and with greater precision than a paper-based tests (Brown 1997 cited in Roever, 2001; Wainer, 2000). Computers and devices such as mobile phones, PDAs, palmtops, smartphones, tablets and netbooks are becoming important resources for learners and are now widely used in language education (Nikou & Economides, 2013).

Besides the traditional paper-based test, there are several types of assessment using technology to administer. For example, in computer-based test, assessment can be delivered in a standalone computer or a closed network computer (Roever, 2001). In addition to computer-based test, web-based test is an assessment that can be delivered via the internet (Nikou & Economides, 2013). The test takers can access the tests by making use of web-browser software to login to the system. Moreover, with the advance of technology, computerized adaptive testing is another way of testing. In other words, computer can select and present test items to the test takers according to the estimated level of their language ability (Dunkel, 1999).

Using technology in assessment can provides many potential advantages including decreasing scoring errors, increasing opportunity to process data online, and the improvement of reliability (Newton et al., 2013). In terms of test administration, testers can administer assessments to several participants at one time, resulting in time, and saving on paper or personnel resources. In comparison with the traditional paper-based test, using technology in assessment can enable the use of multimedia features, which are not workable in the paper-based test format (Kalogeropoulos et al., 2013). In addition, computer-based test and web-based test can provide feedback both on each test takers' response or on the test result immediately upon completion of the test. Moreover, using technology in testing can randomize choices from large questions pool, innovate item formats, automate scoring and reporting, and enhance advanced security (Roever, 2001; Nikou & Economides, 2013).

According to the advantages of using computer technology for educational assessment, there are great transformations from the traditional mode of assessment to the modern method. Many institutions are migrating toward the use of computer-based test. Therefore, they have changed their test platform from traditional paper-based test to technology-based test. For example, the Test of English as a Foreign Language (TOEFL) has changed its test format to the Internet-based test (Coniam, 2006). The University of Cambridge Local Examination Syndicate's (UCLES) CommuniCAT Tests and the International English Language Testing System Test (IELTS) have been changed into computerized format (Coniam, 2006; Hosseini et al., 2014). In addition, computer-based test is also used for many

Veridian E-Journal, Silpakorn University
ISSN 1906 – 3431

examinations in many areas such as educations, militaries, careers and licensing programs (Russo, 2000; Trotter, 2001).

However, there are concerns that should be considered when assessing language skills by using technology-based methods. Several factors have been found to affect performance and lead to the difficulties when taking the tests such as equivalence of scores across methods of test administration or level of computer experience and familiarity with technology of the test takers (Hosseini et al., 2014). Factors originated from test takers themselves such as different styles of learning, individual way of collecting, organizing and processing information, and individuals' attitudes towards using computer and technology have also been argued to influence test performance (Chan, 2013; Newton et al., 2013).

The findings from past research on equivalence of scores across methods of test administration are still inconclusive. For example, it was found that for a dental hygiene course unit in midterm examination, the computer-based tests scores were greater than those of paper-based tests. (DeAngelis, 2000), On the contrary, Gwaltneey et al. (2008) reported lower scores on the computer-based tests version of their patient outcome measures compared to the paper-based tests version. Some studies reported non-significant differences between computer-based and paper-based tests result therefore computer-based tests can be considered a valid and acceptable testing mode (Fyfe et al., 2009; Mason et al., 2001; Schaeffer et al., 1993).

Computer experience and familiarity in the use of computer are also important predictors of learners' achievement in online learning and testing. Younger learners may feel more comfortable learning from technology-based environment as opposed to sitting still listening to a teacher for an hour or more (Lam, 2009). Many researchers have already investigated the relationship between computer usage ability and computer-based test result. Yurdabakan (2012) reports that students with poor computer skills show low achievement in computer-based test. However, Boo (1997) found that there was no significant relationship between computer familiarity and the test takers' performance on three computerized tests. Moreover, Taylor et al. (1999) found no evidence of an undesirable effect of computer familiarity on students' performance on computer-based tests.

According to the policy of The Office of the Basic Education Commission (OBEC) to reform English teaching in schools, students and teachers are encouraged to use technology as tools for language development and assessment (OBEC, 2013). However, the introduction of computer-based and web-based learning and testing may have caused certain difficulties among teachers and students due to their unfamiliarity with computer-based and web-based

learning and testing style, as well as some unexpected technological issues. Therefore, computer-based and web-based resources may be not commonly used in Thai schools even in schools where technology facilities are readily available. Moreover, it is necessary for students to prepare listening as the basic skills for their further study and career in which English plays a vital role in most workplaces (Pratoomrat & Rajprasit, 2014).

Although there is a large body of comparative studies on the paper-based and web-based testing, most studies were conducted in other countries. Particularly on listening performance, there is relatively little research of comparative studies on the paper-based and web-based test performance in Thai context. Moreover, little research has been conducted to investigate the interaction between the test delivery modes and test takers' preference. As discussed above, the comparisons of using technology in testing are still inconclusive. Some studies have reported a significant difference between the two testing modes while others researchers have concluded the opposite (DeAngelis, 2000; Fyfe et al., 2009; Gwaltneey et al., 2008; Mason et al., 2001; Schaeffer et al., 1993). Accordingly, it is necessary for this research to investigate whether the use of traditional paper-based testing and web-based testing to test listening ability would affect test takers' performance or not.

Objectives

This research aimed to compare the test takers' performance on listening test conducted via paper-based and web-based methods and to identify any relationship between the test takers' test mode preference and their performance on listening test.

Research Questions

The study was designed to answer the following research questions.

1. Did test takers perform significantly different on paper-based and web-based listening tests

2. Were there any relationships between the test takers' preference for testing mode and their performance on listening tests?

Materials and Methods

1. Participants

The participants were 200 grade 11 students, 53 males (26.50%) and 147 females (73.50%), ages 16-18. They were studying in the 2015 academic year at Satri Phatthalung School, in the center of Phatthalung Province, South of Thailand.

2. Research Instruments

2.1 Listening Test

Two parallel Cambridge Preliminary English Tests (PET) were used to assess the participants' listening performance. One was administered via paper-based method and the other one was administered via web-based method. Each of the two tests lasted about 30 minutes. The tests consisted of 25 items of multiple-choice, gap-filling and true or false formats.

2.2 Questionnaire

The questionnaire, adapted from the Computer Attitude Scale (CAS) developed by Loyd & Gressard (1984) was used to collect information about the test takers' general information including age, gender and preference for listening test mode. The questionnaire was administered to the test takers after they took the two different listening test modes.

3. Data Collection

The study was conducted in the second semester of the 2015 academic year. The paper-based listening test was administered first to the 200 test takers via an audio player in a normal classroom. Then, in the following week, the participants took the web-based listening test through the learning management system (LMS). The test takers could access this test after logging into the system and listened to the questions via headphones individually. To answer gap filling question types, the test takers needed to type their answers using a keyboard into the gap provided on the screen. For multiple choice and true/false question types, they needed to click their mouse to choose the correct answers. The questionnaire was administered to the participants after the completion of their web-based listening test. All questions for each part in the questionnaire were explained to the participants so they could complete the questionnaire. The time of responding to the questionnaire was about 5 minutes.

4. Data Analysis

Descriptive statistics of paired sample t-tests were used to describe the test takers' performance of their two listening tests: paper-based and web-based. Then, independent sample t-test analysis was used to identify the statistics between test takers' test mode preference and their performance on the listening test. Point biserial correlation coefficient was also used to analyze the relationship between paper-based and web-based score with the variable of participants' test mode preference.

Results and Discussion

This section contains the results of the two mode delivery of listening tests, together with information about participations' preferences on listening test mode and their listening performance.

Research Question 1: Did test takers perform significantly different on paper-based and web-based listening tests?

To answer the first research question, two parallel listening tests were administered to the test takers and Table 1 shows the listening performance of 200 participants on paper-based and web-based methods.

Table 1. Participants' performance on PBT and WBT

Test Mode	Total Score	Mean	S.D.	t	p-value
PBT	25	10.67	2.97	.451	.652
WBT	25	10.57	2.92		

The participants' average scores on the two tests were compared. From the total score of 25, the test takers' mean score on paper-based test (PBT) was 10.67 while that on web-based test (WBT) was 10.57. The results of an analysis showed a non-significant difference between the test takers' mean scores on the two tests. Therefore, it can be concluded that the participants' performance on the paper-based and web-based listening tests were not statistically different. That is, the test mode did not have any effect on the participants' performance. It should be noted that the average scores on both PBT and WBT were less than half, indicating that the participants of the study had low listening proficiency.

Detailed analysis was performed to see if there was any significant difference on the performance of the participants with different listening proficiency. The test takers were divided into high and low proficiency groups based on their scores of paper-based listening test, which was the traditional test mode delivery they were familiar with. Results are presented in Table 2, with test takers divided into high and low group using 27% formula. Both the high and low groups consisted of 54 test takers.

Table 2. Low and high-proficiency participants and their listening test scores

Group	Test Modes	Total Score	Mean	S.D.	t	p-value
High	**PBT**	25	14.46	1.98	**4.243****	0.01
	WBT	25	12.54	3.27		
Low	**PBT**	25	7.33	1.13	**-5.379****	0.01
	WBT	25	9.20	2.41		

** significant at the 0.01 level

Veridian E-Journal, Silpakorn University
ISSN 1906 – 3431

Tables 2 demonstrates that from the total of 25, the high group's mean scores on paper-based test were significantly higher than that on web-based test ($\bar{X}$ =14.46 and $\bar{X}$ =12.54 respectively, P<0.01). In other words, the high group performed significantly better on paper-based test than on web-based test. On the other hand, in the low group, from the total of 25, their mean scores on paper-based test was significantly lower than that on web-based test ($\bar{X}$ =7.33 and $\bar{X}$ =9.20 respectively, P<0.01). That is, the low group performed significantly better on web-based test than on paper-based test.

Research Question 2: Were there any relationships between the test takers' preference for testing mode and their performance on listening tests?

Of the 199 test takers who completed the questionnaire, 33 (16.58%) preferred paper-based method while 166 (83.42%) preferred web-based method. Independent sample t-test was used to describe the statistics of their listening performance. Details of their performance are presented in Table 3.

Table 3. Listening performance by test mode preference

PBT Preference (n=33)				t	Sig. (2-tailed)	WBT Preference (n=166)				t	Sig. (2-tailed)
PBT		WBT				PBT		WBT			
Mean	S.D.	Mean	S.D.	3.59**	.000	Mean	S.D.	Mean	S.D.	-.232	.816
12.30	3.046	10.33	2.859			10.45	3.260	10.58	2.861		

** significant at the 0.01 level

Table 3 shows that 33 test takers who were in favor of paper-based listening test performed significantly better on paper-based test ($\bar{X}$ =12.30) than on the web-based test ($\bar{X}$ =10.33), indicating that their preference on paper-based method helped them to perform well on paper-based testing. Of 166 test takers who preferred web-based method, their mean scores on web-based ($\bar{X}$ =10.58) and paper-based test ($\bar{X}$ =10.45) were not significantly different.

Point biserial correlation coefficient was used to see if there was any correlation between the participants' test mode preference and their performance.

Table 4. Correlation of scores and test mode preference

Test mode preference	PBT Scores	WBT Scores
	r	r
PBT Preference	**.248****	
WEB Preference		.017

** significant at the 0.01 level

As can be seen from Table 4, the overall correlation of scores and test channel preference of test takers who preferred paper-based test was .248, a low but significant correlation. However, for those who preferred the web-based test mode, the correlation was .017 which indicated that there was no association between the test mode preference and their WBT scores.

Point biserial correlation coefficient was also used to investigate if there was any correlation between the participants' test mode preference and their performance based on the level of listening proficiency.

Table 5. Correlation of scores and test channel preference of high and low groups

Group	Test mode preference	PBT Score	WBT Score
		r	r
High	PBT Preference	.034	
	WEB Preference		.219
Low	PBT Preference	.117	
	WEB Preference		**.263***

* significant at the 0.01 level

Surprisingly, as illustrated in Table 5, more detailed analysis for each group of test takers showed that there was only a significant correlation between the test mode preference and the web-based listening test scores in the low group. However, the level of the correlation was very low, indicating that their preference for web-based test method might somehow enabled them to do better on web-based listening test.

Conclusion

The study revealed that there were no significant differences between 200 participants' performance on the two types of listening test mode delivery. In other words, difference in test mode was not found to be a factor affecting the test performance of overall test takers. The finding was found to be similar to some previous studies in that there was no significant difference between computer-based and paper-based tests result (Fyfe et al., 2009;

Veridian E-Journal, Silpakorn University
ISSN 1906 – 3431

Mason et al., 2001; Schaeffer et al., 1993). Thus, instructors may include web-based listening test method in their listening tests without affecting the students' test performance.

However, detailed analysis revealed that the high-proficiency group performed significantly better on paper-based test than on web-based test while the low-proficiency group performed significantly better on web-based test than on paper-based test. The findings of an analysis seem to suggest that listening proficiency levels might have played an important role in the effect of test mode delivery. Teachers and test developers may consider using both test modes simultaneously in order to administer the reliable test with test takers who are different in listening proficiency.

In terms of the relationship between test mode preference and test performance, it was found that test takers in favor of paper-based method performed significantly better on paper-based test than on web-based test whereas the web-based scores of those who preferred web-based method were not significantly higher than their paper-based scores. It was interesting to find that the familiarity with traditional paper-based listening test might help paper-based group to do better on their preferred test mode. In addition, no relationship was found in detailed analysis between mode preference and test performance of the high-proficiency group while there was a significant relation between the performance of the low-proficiency group. That is, low proficiency test takers in favor of web-based mode performed significantly better on web-based test than on paper-based test while those who preferred paper-based method did not performed better on paper-based test.

The findings of the current study may help the instructors and test developers to realize the effect of test mode preference on test performance. The fact that 166 out of the 200 participants preferred web-based listening tests would also be useful information for language teachers, educators, test developers, and test administrators to provide appropriate testing modes to test their students' listening ability. As technology has been implemented in the field of language assessment, if students are well prepared for the technology-based exams, their performance will be enhanced. This finding can also be a beneficial information for the administrator to set the policy such as Investing in test management software or training and supporting unit, as well as improving the technological awareness and utilization of technology-based test at the primary and secondary school level for both the teachers and the students.

Further research needs to be conducted before any conclusion of the effects of test mode delivery and test mode preference on test performance can be drawn. The result implies the need of further studies on other factors that may influence the test performance such as computer experience, familiarity with technology, learning styles, or attitudes towards using computer and internet. Moreover, the influence of using technology in testing listening skills is needed to examine on how technology might affect listening performance, particularly, the possibility that technology might somehow distort or impede the test performance.

References

Boo, J. (1997). **Computerized versus paper-and-pencil assessment of educational development: Score comparability and examinee preferences**. Unpublished PhD dissertation, University of Iowa, USA.

Byrnes, H. (1984). The role of comprehension: A theoretical base. **Foreign Language Annuals**, 17, 317-329.

Chan, P. M. (2013). Learning styles. **วารสารวิชาการ Veridian E-Journal**, 6(7), 39-41.

Coakley, C., and Wolvin, A. (1997). Listening in the educational environment. In M. Purdy & D. Borisoff (Eds.), **Listening in everyday life: A personal and professional approach** (2nd ed.) (pp. 179-212). Lanham, MD: University Press of America.

Coniam, D. (2006). Evaluating computer-based and paper-based versions of an English language listening test. **ReCALL**, 18, 193-211.

DeAngelis, S. (2000). Equivalency of computer-based and paper-and-pencil testing. **Journal of Allied Health**, 29, 161-164.

Dunkel, P. A. (1999). Considerations in developing and using computer-adaptive tests to assess second language proficiency. **Language learning & technology**, 2(2), 77-93.

Fyfe G., Meyer J., Fyfe S., Ziman M., Sanders K., and Hill J. (2009). Self- evaluation of assessment performance can enhance student's perception of feedback on computer-generated tests. Conference: **Proceedings of the 35th International Association for Educational Assessment Annual Conference**, Brisbane, Australia.

Gilbert, M. B. (1988). Listening in school: I know you can hear me--But are you listening? **Journal of the International Listening Association**, 2, 121-132.

Gwaltney, C. J., Shields, A. L., and Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. **Value in Health**, 11, 322–333.

Hosseini, M., Abidin, M. Z., and 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.

Kalogeropoulos, N., Tzigounakis, I., Pavlatou, E., and Boudouvis, A. (2013). Computer-based assessment of student performance in programming courses. **Computer applications in engineering education**, 21(4), 671–683.

Lam, M. (2009). Effectiveness of web-based courses on technical learning. **Journal of Education for Business**, 84(6), 323–331.

Loyd, B., and Gressard, C. (1984). Reliability and factorial validity of computer attitude scales. **Educational and psychological measurement**, 44(2), 501-505.

Mason, B. J., Patry, M., and Bernstein, D. J. (2001). An examination of the equivalence between non-adaptive computer-based and traditional testing. **Journal of Educational computing research**. 24, 29-39.

Mendelsohn, D. (1994). **Learning to listen: A strategy-based approach for the second language learner**. California, USA: Dominie Press.

Newton, C., Acres, K., and Bruce, C. (2013). A comparison of computerized and paper-based language tests with adults with aphasia. **American journal of speech-language pathology**,22, 185-197.

Nikou, S., and Economides, A. A. (2013). Student achievement in paper, computer/web and mobile based assessment. **BCI**, (pp. 107-114).

Nunan, D. (1998). Approaches to teaching listening in language classroom. In **Proceedings of the 1997 Korea TESOL Conference**. Taejon, Korea: KOTESOL.

OBEC. (2013). Practicality of ministry of education policy on reforming English teaching for school. Office of The Basic Education Commission, Ministry of Education.

Pratoomrat, P., and Rajprasit, K. (2014). Exploring current situations and corporate needs of English language use in workplace: Thai professionals' voices to tertiary education. **วารสารวิชาการ Veridian E-Journal**, 7(4), 28-47.

Roever, C. (2001). Web-based language testing. **Language learning & technology**, 5(2), 84-94.

Russo, A. (2000). **Mixing technology and testing**. Retrieved from The School Administrator (Online): http://www.aasa.org/publicstons/sa/2002_04/Russo.htm

Schaeffer, G., Reese, C., Steffen, M. McKinley, R. and Mills, C. (1993). Field Test of a Computer-Based GRE General Test. **Reports-Research/Technical**. ETS-RR-93-07.

Shang, H. (2008). Listening strategy use and linguistic patterns in listening comprehension by EFL learners. **The international journal of listening**, 22, 29-45.

Taylor, C., Kirsch, I., Eignor, D., and Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer- based language tasks. **Language Learning**, 49(2), 219-274.

Thomas, I., and Dyer, B. (2007). **The problem of poor listening skills**. Utah: Weber State University.

Trotter, A. (2001). Testing firms see future market in online assessment. **Education week on the web**, 20(4), 6.

Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. **ELT Journal**, 53(3), 168-176.

Wainer, H. (2000). **Computerized adaptive testing: A primer**. New York: Routledge.

Waye, K. P., Magnusson, L., Fredriksson, S., and Croy, I. (2015). A screening approach for classroom acoustics using web-based listening tests and subjective ratings. **PLOS ONE**, 10(1).

Wolvin, A. D., and Coakley, C. G. (1988). **Listening (3rd ed.)**. Dubuque, IA: Wm. C. Brown

Wu, W. S. (2005). Research and development of online adaptive placement test of listening comprehension: A pre-liminary report, **Journal of Education and Foreign Languages and Literature**, 1, 147-160.

Yurdabakan, I. (2012). Primary school students' attitudes towards computer based testing and assessment in Turkey. **Turkish Online Journal of Distance Education**, 13 (12), 177-188.