

Iranian Attitudes and Test Mode Preference: Factors Contributing to the Acceptance of Computerized Test

Jafre Abidin & Monirosadat Husseini

*Universiti Sains Malaysia
11800 Penang, Malaysia
email: ho.mahmonir@yahoo.com*

Abstract

This article explores the probable relationship between the attitude towards the use of computer and test performance on Computer-Based-Tests (CBT) among Iranian university English as foreign language (EFL) learners and comparing the results of CBT and Paper-and-Pencil Test (PPT). Analyzing data from test results and follow up interviews, this study revealed that test takers showed a high positive attitude and preference towards the use of computer and computerized tests. This study adapted a mixed-method approach. A sample of 202 Iranian EFL learners participated in the study given two equivalent English proficiency tests in computerized and paper and pencil types. Using paired-sample t-test, correlational analysis, and ANOVA, the findings of this study indicated that Iranian test takers had overall positive attitudes towards computerized tests with higher scores on paper and pencil tests. The qualitative analysis showed that they mostly preferred computerized test believing that it largely reflected their overall English knowledge. In addition, it was found that test takers did better on PPT because they believed that this type of testing was more familiar and less threatening for students. The findings of this study have implications for test developers and curriculum providers in the light of further improvement and promotion of test and the process of test administration.

Keywords *Test taker's attitude, Test mode preference, Computer-Based &Test, Paper-and-Pencil Test, EFL learners*

INTRODUCTION

The necessity of using technological devices in teaching and testing English language in educational settings has been rapidly increasing since broad developments in information and communication technology (Bachman, 2000; Chapelle, 2007). The widespread use of computer technology in conducting language tests and the availability of personal computers, along with increased computer familiarity are making computer-based test administration feasible on a large scale. In this regards, computer based testing is going to be applied all around the world in academic contexts (Chappel & Douglas, 2006; Clarian & Wallace, 2002). Given the integral role computers play in our lives, the number of computer-delivered tests is increasing in language assessment due to the perceived advantages of computer-delivered tests. Since 1990s, many researchers advocated the importance of assessment in helping students learn better. Earl (2003) describes examinations in today's schools as primarily evaluation of learning. On the other hand, computerized testing advocates believed that traditional measurement implementation place too much emphasis on passing a test rather than on encouraging learners to learn beyond education (Tanner, 2001).

In addition, in language learning, as the use of computers and electronic devices has become popular; especially in assessing the English learners' language proficiency, the most precise and available way could be through computers (Bachan, 2000; Clarian & Wallace, 2002; Paek, 2005; Sawaki, 2001). However, the high cost of using computers in high stake tests and less computer familiarity had limited the implementation of computerized language tests (Clarian & Wallace, 2002). Bugbee and Brent (1990) asserted that once set up, Computer-Based-Tests (CBT) is easier to administer than Paper-and-Pencil Test (PPT) and offer the possibility of instant grading. Fan (2014) suggested that in computerized examination, testing conditions can be standardized and the sequence of items can be easily manipulated.

REVIEW OF LITERATURE

Testing is just one form of assessment and means gathering information about a particular learner's behavior in various forms (Hughes, 2003). In language learning settings, information about people's language ability is often very useful and somehow necessary. For example, accepting students from overseas or hiring interpreters or translators in an organization without some idea about their proficiency of English is difficult to imagine. The same is true in the judgment about the students going to upper levels of their study. They certainly need dependable measures of language ability. Test results are also needed to provide information about the achievement of learners, without which it is difficult to see how rationale educational decisions can be made. In this regards, Hughes (2003) states that, "while teachers' informal assessments of their students for some purposes are both appropriate and sufficient, this is not true for the cases just mentioned. Even without considering the possibility of bias, we have to recognize the need for a common yardstick, which tests provide, in order to make meaningful comparisons" (p. 4).

Bachman and Palmer (2000) defined language test performance as a particular instance of language use in specific situations. They claim that if instructors want to use the test results as a judge about learner's language ability and make some decisions, they need a conceptual framework to be able to demonstrate how performance on a particular language test is related to language use in specific situations other than the language itself. It means they need a framework that enables them to use the same characteristics to describe what they believe are the critical features of both language test performance and non-test language use.

The characteristic of individuals that is of primary interest in language testing is language ability, which is what the educators intend to make inferences about. Other individual characteristics that are needed to be considered are topical knowledge of knowledge schemata. These are important because these characteristics can have important influence on both language use and test performance; in addition, it is possible to design language tests so that these characteristics facilitate, rather than impede, test takers' performance. Figure 1 illustrates the relationship between language use and language test performance and language users' characteristics provided by Bachman and Palmer (2000).

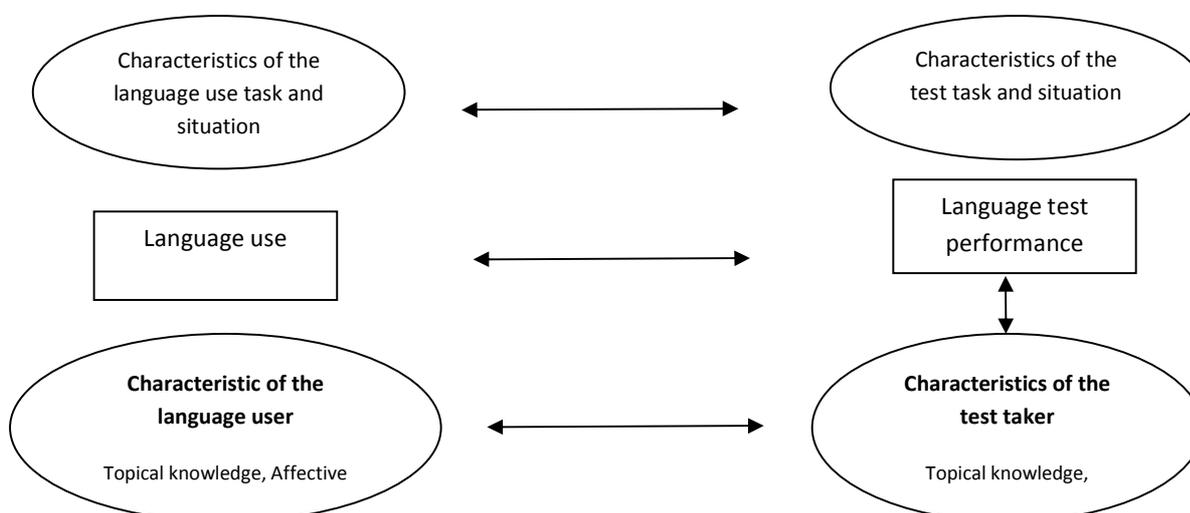


Figure 1 Correspondence between language use and language test performance (Bachman & Palmer, 2000)

There are various personal characteristics that test developers should consider while designing or developing tests in second or foreign language learning settings. "Personal characteristics are individual attributes that are not part of test takers' language ability but may still influence their performance on

language test". (Bachman & Palmer, 2000, p. 12). Bachman and Palmer (2000) pointed out that in any test development project, the test designer needs to provide a specific list of personal characteristics that are potentially contributed to the usefulness of test. For example, information on some personal characteristics, such as age, gender, first language ability, level of general education, or type and amount of prior experience with a particular test may have a significant influence on the qualities of usefulness, and collecting information on these personal characteristics may be quite valuable in designing test tasks to optimize the qualities.

McNamara cited in Davies and Elder (2004) describes an examination as a process of eliciting information about respondents' ability from observed performance, under particular test conditions, in order to draw inferences about the likely quality of their performance under non-test conditions or their situation in relation to a relevant domain of abilities. Then he makes a distinction between the test (the means of drawing inferences) and the criterion (the target of test inferences) and illustrates a relation of test, construct, and criterion in a model (figure 2).

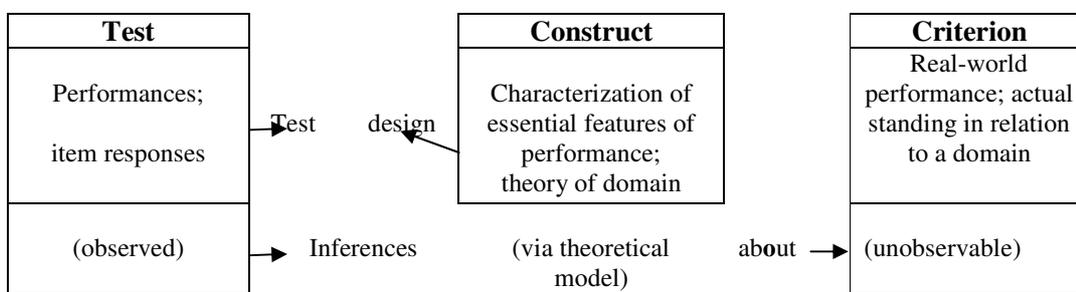


Figure 2 Test, construct, and criterion (Davies and Elder, 2004)

According to this definition, a language test is used to predict the student's language proficiency to cope with the prerequisites of the university settings. These demands need considerable research to find out what is required of English students in university settings in both academic and social domains through a model. This information, then, is used to build up a picture of academic language proficiency, and hence it is reflected in the design of the test. Such a construct may be matched, for example, in the degree to which the demands of specific areas of study are included, or ignored (McNamara cited in Davies and Elder, 2004). This model is used because there is a strong relationship between test designs (formats), construct (test takers performance and ability), and the criterion (students standing in university contexts as distance educational systems).

Comparability Model

In reviewing the previous studies on the comparability of PPT and CBT, neither of them developed its own model to scrutinize comparability, or adapted a model from the literature despite the scarcity of such models. To the best knowledge of the researcher, there is only one model presented by McDonald (2002) which demonstrates the testing experience under two test modes, and how it affects the statistical equivalence of two testing modes. McDonald's model is presented in figure 3.

McDonald, in his model, illustrates the interaction between the test-taker and the test where the testing mode is different. The results is what Honaker (1988, p. 562) called 'experiential equivalence', i.e. "emotional, perceptual, and attitudinal reactions to the two forms". The model shows that test-taker experience is influenced by two factors: the interaction between the test-taker and the testing mode, and the differences between the two testing modes.

McDonald (2002) believed that individual characteristics such as computer familiarity, computer

anxiety, and computer attitudes could contribute to the reaction of test-takers towards the testing mode. In addition, he pointed out that test-takers are also influenced by test contexts, text presentation, reviewing or ignoring some questions which may in turn influence test-takers' performance. As a result, the test takers' experience from two different test contexts, and therefore, the construct being measured could be affected differently, which is shown in the results of statistical comparability of two testing modes.

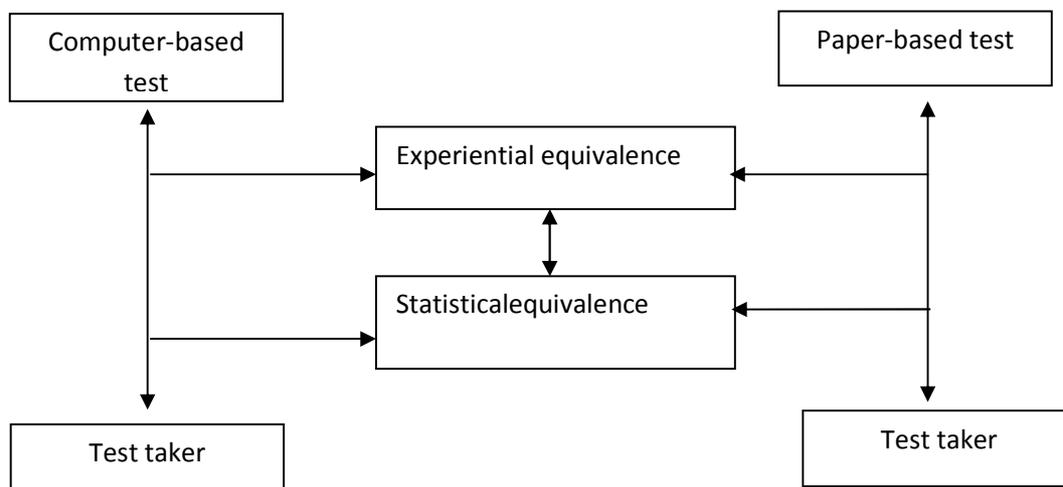


Figure 3 McDonald's (2002) comparability Model

On the other hand, attitude, defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly&Shelly, 1998, p. 1) has been extensively researched in social psychology. In the field of second language acquisition, a learner's positive attitudes towards the target language or the target-language culture or method of teaching have a beneficial impact on their language learning (e.g. Fan, 2014). Gardner (1985), for example, believes that attitude and other affective variables are as important as aptitude for language achievement.

Despite the widely recognized importance of attitude in second language learning, there is no evidence that it is adequately investigated in the field of language testing (Fan, 2014). Based on TAM theory (Davis, 1989), not only having positive attitude towards the use of computer could influence the performance on English language computerized tests, but also the good performance on tests also, by itself, could affect test taker's attitude towards the use of computers which in turn could affect the actual use of technology in language learning and testing. A test taker's attitude towards a language test forms part of test impact, which is one of the essential qualities in Bachman and Palmer's (2000) test usefulness framework and described by Shohamy (2001) as consequential validity. Messick (1989) also explicitly recommends that attitude be considered as a crucial source of evidence for construct validity. Therefore, as test takers' attitude is believed to be part of test impact, a fair test should function equally among different groups of test takers.

Davis (1989) believes that perceived usefulness and perceived ease of use, according to TAM theory, have a significant influence on a user's attitude towards using a system. Davis (1989) defined usefulness as the degree to which an individual sees using computer will enhance his performance, and perceived ease of use as the degree to which using computer will be free of effort. Davis (1989) then proposed that behavioral intention to use the system, here computer in test, is modeled as a function of attitude towards the use of computer and usefulness, which in turn determines the actual use of the individual. According to this model, the feeling of usefulness and ease of use influence the attitude of the user. Then the positive or negative attitude may affect the intention of the individual in using the system.

This behavioral intention, determines the actual use of the system (in this study test performance is used) properly or improperly.

The present study aims at investigating the relationship between test takers' attitude towards computerized tests and test performance in computer-based tests (CBT). In addition, it probes the comparability between CBT and PPT among 202 Iranian EFL learners by comparing two test results. The same group was given two equivalent tests on CBT and PPT.

Purpose of the Study

In the light of research literature on importance of test takers' attitude towards using computer in testing English language, the main purpose of this study was to find out Iranian EFL students' attitude towards computer based test and then explore the relationship between their attitudes and their test scores on CBT. It also attempts to explore the preference of test takers on CBT and PPT as well as comparing their test scores on both test modes.

Research Questions

Specifically, the research questions are:

1. Is there any significant difference between computer-based tests and paper-and-paper based tests when assessing General English among Iranian EFL students?
2. What is the test takers' attitude towards computer-based tests?
3. Is there any relationship between test takers' attitude towards taking computer tests and their test scores in CBT and PPT?

METHOD

Participants

The target population of the study was all English language learners of Tehran Payame Noor University who have registered on General English Language course in first semester of 2013-2014. Among them, 202 were selected randomly as participants of the study.

Instruments

As the purpose of the study was comparing test results on CBT and PPT, two different but equivalent tests elicited from the course book were developed and were validated by three English language teachers and piloted by 30 similar students to test their validity and reliability. As one of the main variables of study was attitude towards the use of computer, the current study adapted the most popular instrument called Computer Attitude Scale (CAS) developed by Loyd and Gressard (1985), consisting of four components of computer anxiety, computer confidence, computer liking, and computer usefulness and a whole measure of 40 items. The questionnaire is a five Likert scale with the score of 5 for high attitude and one for the low attitude towards the use of computer. To ensure its reliability, the Cronbach's alpha reliability coefficient was calculated. Based on Cronbach's formula, the researchers achieved an alpha value of 0.91, which indicates a very acceptable level of reliability for the purpose of this study. In addition, a semi structure interview was administered to elicit information about their preference for CBT and PPT and their reasons for their answers.

Procedure

The process of data collection for this study was done during September to December of 2013. Students were given paper-based test first after two months of the beginning of the course and after three weeks the equivalent computerized version was administered among the same group. At the end of CBT, the students were given an additional 15 minutes to complete the attitude questionnaire. After conducting both tests, the researcher has done the interview among 8 students to know their preference for test modes of CBT and PPT and their explanation for their choices. Efforts to compare the differences in the test takers' scores between CBT and PPT developed the first null hypothesis:

H1₀: There is no significant difference between test scores of CBT and PPT among Iranian EFL students.

The hypothesis was tested using paired sample t-test to find out the probable difference between two tests. Analysis of variance allows the existing or not existing possible difference and the type of difference, whether significant or not. It could be justified that if there is a learning curve in the case of existing difference between two test mediums, an interaction would be present as the scores for subsequent tests would become increasingly different. To see whether there is significant relationship between attitude and test scores and if the test medium had any interaction between CBT and PPT, Pearson Correlation Coefficient and within and between subjects ANOVA were run to test the second and third null hypotheses:

H2₀: There is no significant relationship between PNU English language learners' attitudes towards the use of computer and their test result on CBT.

H3₀: There is no significant difference between attitude towards the use of computer and traditional test mode and test results on CBT.

In order to function properly, the ANOVA program required an equal number of data points for both test scores; in addition, in order to ensure the accuracy when considering interaction between two sets of scores, it was desired that for each subject represented in the analysis, all of that subjects respective test scores fell consistently in the same population. All of the students who took the paper-based test were participated in computer-based tests.

RESULTS

In order to test the first null hypothesis, a paired-sample t-test at 0.05 levels of significance was conducted and the frequencies of each item in the questionnaire were calculated respectively. The descriptive statistics for PPT and CBT is depicted in Table 1.

Table 1 Descriptive Statistics of CBT and PPT Results

	Mean	N	Std. Deviation	Std. Error Mean
Scores on PPT	14.69	202	3.14	.22
Scores on CBT	13.24	202	2.31	.16

According to the Table, the students showed a higher performance on PPT (M = 14.69, SD = 3.14) than CBT (M = 13.24, SD = 2.31). Table 2 presents the result of Paired-Sample t-test.

The results of the paired-sample t-test on Table 2 ($t(201) = 6.72$, $P < .05$) indicated that there is a significant difference between the subjects' means on CBT and PPT scores. The subjects performed better on PPT. Thus, the first null-hypothesis is rejected, that is there is significant difference in test results of two tests with the priority of PPT over CBT.

Table 2 Paired Samples t-test of PPT and CBT Scores

Statistics	Mean	Std.	Std. Error Mean	95% Confidence Interval of the Differences		t	Df	Sig. (2-Tailed)
				Lower	Upper			
	1.45	3.06	.21	1.02	1.87	6.72	201	.000

To analyze the questionnaire, the researcher made use of only the simple comparisons of frequency percentages related to the predefined answers for the questionnaire's questions, namely, Strongly Agree, Disagree, Neutral, Disagree, and Strongly Disagree. This measure identified which item has the most frequency among the others and consequently determined the degree of students' attitude towards CBT. Table 3 and 4 depict the descriptive statistics and the column chart associated with the questionnaire respectively.

Table 3 Descriptive Statistics of Attitude towards the Use of Computer

	N	Minimum	Maximum	Mean	Std. Deviation
ATT	202	2.03	5.00	3.77	3.52
Valid N (listwise)	202				

Table 3 indicates that the mean of attitude towards the use of computer is high ($M = 3.77$, $SD = 3.52$). Table 4 indicates the descriptive statistics of components of questionnaire of attitude. The criteria utilized for each item were as follows: 5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, and 1 = strongly disagree. As it can be seen in Table 4, Items 10 to 29, the participants selected 'Agree' with the highest percentage and the second rate of high percentage belongs to 'Strongly Agree'. Besides, the items 1-9 for anxiety, the high percentage belong to 'Strongly Disagree' and 'Disagree'. That is the majority of students showed low anxiety and high confidence, liking, and usefulness towards using computer. In general, they reported high positive attitude towards the use of computer.

Table 4 Results of CAS: Participants' Attitude towards the Use of Computer

Items	Statements	SD		D		N		A		SA		Mean
		F	P	F	P	F	P	F	P	F	P	
*1	Anxiety 1	78	38.6	66	32.7	30	14.9	25	12.4	3	1.5	2.05
*2	Anxiety 2	73	36.1	84	41.6	23	11.4	21	10.4	1	5	1.97
*3	Anxiety 3	72	35.6	75	37.1	23	11.4	27	13.4	5	2.5	2.09
*4	Anxiety 4	99	49	73	36.1	20	9.9	9	4.5	1	.5	1.71
*5	Anxiety 6	84	41.6	79	39.1	24	11.9	13	6.4	2	1	1.86
*6	Anxiety 7	64	31.7	85	42.1	29	14.4	19	9.4	5	2.5	2.08
*7	Anxiety 8	90	44.6	85	42.1	18	8.9	8	4	1	.5	1.73
*8	Anxiety 9	75	37.1	89	44.1	17	8.4	19	9.4	2	1	1.93
*9	Anxiety 10	82	40.6	78	38.6	23	11.4	15	7.4	4	2	1.91
10	Confidenc1	5	2.5	14	6.9	49	24.3	102	50.5	32	15.8	3.70
11	Confidenc2	5	2.5	16	7.9	41	20.3	111	55	29	14.4	3.70
12	Confidenc3	22	10.9	64	31.7	51	25.2	50	24.8	15	7.4	2.86
13	Confidenc4	8	4	34	16.8	46	22.8	90	44.6	24	11.9	3.43
14	Confidenc6	3	1.5	20	9.9	55	27.2	101	50	23	11.4	3.59
15	Confidenc7	4	2	19	9.4	32	15.8	112	55.4	35	17.3	3.76
16	Confidenc8	12	5.9	33	16.3	74	36.6	63	31.2	20	9.9	3.22
17	Confidenc9	4	2	23	11.4	64	31.7	82	40.6	29	14.4	3.53

18	Like 1	1	.5	8	4	23	11.4	111	55	59	29.2	4.08
19	Like 2	2	1	15	7.4	50	24.8	96	47.5	39	19.3	3.76
20	Like 3	2	1	14	6.9	51	25.2	98	48.5	37	18.3	3.76
21	Like 5	5	2.5	9	4.5	30	14.9	112	55.4	46	22.8	3.91
22	Like 7	19	9.4	60	29.7	49	24.3	54	26.7	20	9.9	2.98
23	Like 8	21	10.4	71	35.1	59	27.7	41	20.3	13	6.4	2.77
24	Like 9	11	5.4	49	24.3	59	29.2	66	32.7	17	8.4	3.14
25	Usefulness4	1	.5	5	2.5	21	10.4	87	43.1	88	43.6	4.27
26	Usefulness5	1	.5	8	4	27	13.4	95	47	71	35.1	4.12
27	Usefulness6	2	1	13	6.4	27	13.4	108	53.5	52	25.7	3.96
28	Usefulness7	2	1	4	2	19	9.4	67	33.2	110	54.5	4.38
29	Usefulness8	1	.5	13	6.4	52	25.7	72	35.6	64	31.7	3.91

SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree, F = frequency, P = percentage

Regarding the first research question, the analysis of the result proved that the kind of administration in CBT and PPT are different with the higher score in PPT than in CBT despite the higher attitude and good feeling towards the use of computer in their exams. Table 5 shows the result of Pearson Correlation Coefficient in analyzing the relationship between attitude scores of participants and their test scores in CBT.

Table 5 Pearson Correlation of Attitude and CBT Scores

		Attitude
CBT	Pearson Correlation	-.080
Score	Sig. (2-tailed)	.259
	N	202

The results ($R(200) = -.080, P > .05$) indicated that there is no significant relationship between the two variables (Table 5). Thus, the second null-hypothesis is correct as there is no significant relationship between PNU English language learners' attitudes towards the use of computer and their test result on CBT.

Table 6 ANOVA for Test of Within-Subjects Effects of Attitude towards PPT and CBT and test scores of CBT

Source		Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Scores	Sphericity Assumed	8096.282	1	8096.282	2782.199	.000	.933
	Greenhouse-Geisser	8096.282	1.000	8096.282	2782.199	.000	.933
	Huynh-Feldt	8096.282	1.000	8096.282	2782.199	.000	.933
	Lower-bound	8096.282	1.000	8096.282	2782.199	.000	.933
Scores *	Sphericity Assumed	1.565	1	1.565	.538	.464	.003
	Greenhouse-Geisser	1.565	1.000	1.565	.538	.464	.003
	Huynh-Feldt	1.565	1.000	1.565	.538	.464	.003
	Lower-bound	1.565	1.000	1.565	.538	.464	.003

To test the third null hypothesis the within-subject ANOVA was utilized for determining the difference, effect, and degree of difference between students attitude towards the computer based and paper based tests and their test scores on CBT. Table 6 shows the result of analysis.

The results of Table 6 show the significance value is .000. It is less than 0.05 that means there is significant difference between the language learners' attitude and CBT scores with the effect size of .933, which according to Cohen (1988) is very large effect size. However, the results of between-subject analysis in Table 7 show that there is no significant difference between attitude towards PPT and CBT at the significant level of more than .05.

Table 7 ANOVA for Test of Between-Subjects Effects of PPT and CBT Attitudes

Source	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	1	25977.298	9558.778	.000	.980
attitude	1	.127	.047	.829	.000
Error	200	2.718			

The significant value is 0.829, which is above 0.05 level and indicates non-significant difference in CBT and attitude towards PPT and CBT. So the third null hypothesis is supported in that there is no significant difference between CBT and attitudes towards CBT and PPT.

The result of interview indicated that the students showed more preference for CBT than PPT despite their higher score on PPT. Their justification for this contradiction was the more familiarity with traditional method of testing and habit of doing in this type because of the more frequency of doing it since school and high school periods. In addition, they argued that taking English exams through computer is different with taking other native language tests because of the excitement of using technology in doing exam and remembering English knowledge simultaneously. They mostly complained about using computer in only final exams and not during the courses and believed that it is better to use computer not only for exams, but also in learning contexts during the course and also midterm exams.

DISCUSSION

Regarding the first research question, the analysis of the results showed the significant difference between test scores of CBT and PPT. The finding is in line with Pomplun, Frey and Becker (2002) and Choi, Kim and Boo (2003), who found differences between mean scores with higher performance on PPT. Al-Amri (2008) also did the same comparison study and found differences between two test scores with slightly higher score in PPT. He argued that this difference could be due to the eye fatigue while reading text through screen as well as the novelty of presenting test through computer in his study context. In addition, Flowers, Do-Hong, Lewis, and Davis (2011) in their study on investigating the difference between test scores of students on CBT and PPT found students perform better on PPT. They hypothesized that one reason for the lower score for computer-based test might be due to giving the opportunity of read-aloud accommodation in the traditional mode of test presentation, PPT, which was not given in CBT. The justification of the difference hence back to the strategy students used in finding the answer of questions. In contrast, Clariana and Wallace (2002) and Maguire, Smith, Brallier, and Palm's (2010) study on the comparison between CBT and PPT revealed significant difference between test scores with better performance in computerized test. However, this finding may be at odds with some previous studies. For instance, Anakwe (2008), and Mojarrad, Hemmati, JafariGohar, Sadeghi(2013) did the comparison study on PPT and CBT and found no significant difference between test scores on both test modes.

Regarding the second research question, the analysis of the attitude questionnaire at the time of interview indicated that students had a better sense and preference toward taking the examination using computers. This finding is consistent with the results of Higgins, Russell, and, Hoffmann (2005) who found that their participants have preferred to take the test on computer. Likewise, Way, Davis, and Fitzpatrick (2006) found that students feel comfortable with taking tests on the computer and prefer it to paper tests. In addition, Chua (2012) showed that doing computer based tests increased test takers'

curiosity, self-efficacy, involvement, joyfulness, comprehension, and social dimensions and hence students prefer CBT to PPT.

In contrast, the results do not agree with the result of some studies such as Ward, Hooper, and Hannafin (1989) that indicated a negative attitude towards computer testing. Moreover, Sam, Othman, and Nordin (2005) found that the participants showed medium and not high attitude towards using the Internet.

Regarding test mode preference, the overall findings in the related literature show that students preferred computer-based to paper-and-pencil testing (Al-Amri, 2009; Flower et al., 2011; Higgins et al., 2005; Mojarrad et al., 2013; Yurdabakan&Uzunkavan, 2012) while very few studies showed students' preference for PPT over CBT (Harde, Crowson, Xie, & Ly, 2006).

Noyes, Garland, and Robbins(2004) in justifying the contrast between the preference on CBT and better performance on PPT stated that students' task demand becomes higher when they encounter with computerized tests. Li and Pu (2010) found that students who used computer in learning during the course, experienced lower workload than those who did not experience using computer in learning process. The finding of this study revealed that one of the key factors of better performance is related to the habit of doing a particular test type, here PPT. Thus, habit can be regarded as one of the determinant factor influencing the test performance of test takers in PPT compared to CBT.

CONCLUSION

Based on the discussion of the study, it can be concluded that with the high attitude scores towards the use of computer despite the higher score on PPT over CBT, the teachers, schools, universities, and institutes can benefit from technology particularly the computers and internet in their educational environments to promote and encourage students towards assessments. Furthermore, despite one of the goals of PNU as a distance educational system is providing students with the ability to use technology appropriately and adequately, not all students and staff members accept using technology with the same degree of enthusiasm and confidence. This drawback can be minimized by making students more familiar with such environments with the help of technology in learning contexts before encountering with electronic testing.

REFERENCES

- Al-Amir, S. (2009). Computer based testing vs. paper based testing: Establishing the comparability of reading tests through the revolution of a new comparability model in a Saudi EFL context. Unpublished thesis submitted for the degree of Doctor of Philosophy in Linguistics. University of Essex (UK).
- Anakwe, B. (2008). Comparison of student performance in paper-based versus computer-based testing. *Journal of Education for Business, September/October*, 13-17.
- Bachman, L. F., & Palmer, A. S. (2000). *Language Testing in Practice*. (3rded.). Oxford University Press. UK.
- Bugbee, A., & Brent, F. (1990). Testing by computers: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education*, 23(1), 87-100.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. New York: Cambridge University Press.
- Chapelle, C. (2007). Technology and second language acquisition. *Annual Review of Applied Linguistics*. 27, 98-114.
- Choi, I., Kim, K., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295-320.
- Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behaviour*. 28(5), 1580-1586.
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*. 33(5) 593-602.

- Cohen, A. D. (1994). *Assessing Language Ability in the Classroom*. Second edition. New York: Harper and Row, Publishers.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319–340.
- Davies, A., & Elder, C. (2004). *The Handbook of Applied Linguistic*. UK: Blackwell Publishing.
- Eagly, A. H., & Shelly C. (1998). "Attitude Structure and Function". In D.T. Gilbert, Susan T. Fisk, and G. Lindsey (eds.), *Handbook of Social Psychology* (pp. 269–322). New York: McGowan-Hill.
- Earl, L. M. (Ed.). (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Fan J. (2014). Chinese test takers' attitudes towards the VersantEnglish Test: a mixed-methods approach. *Language Testing in Asia*, 4(6), 1-17.
- Flowers, C., Do-Hong, K., Lewis, P., & Davis, V. C. (2011). A comparison of computer-based testing and pencil-and-paper testing for students with a read- aloud accommodation. *Journal of Special Education Technology*, 26(1), 1-12.
- Gardner, R. (1985). *Social Psychology and Second Language Learning: The Role of Attitude & Motivation*. London: EdwardArnold.
- Harde, P. L., Crowson, H. M., Xie, K., & Ly, C. (2007). Testing differential effects of computer-based, web-based, and paper-based administration of questionnaire research instruments. *British Journal of Educational Technology*, 30(1), 5-22.
- Higgins, J., Russell, M., & Hoffmann, T. (2005). Examining the Effect of Computer-Based Passage Presentation on Reading Test Performance. *The Journal of Technology, Learning, and Assessment*, 3(4), pp. 5-34.
- Honaker, L. (1988). The equivalence of computerised and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8, 561-77.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Li, M., & Pu, H. (2010). Comparison between CBT and PBT: assessment of gap-filling and multiple-choice cloze in reading comprehension. *Journal of Language Teaching and Research*, 1(6), 935-941.
- Loyd, B. H., & Gressard, C. (1985). Reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement*, 44, 501-505.
- Maguire, K. A., Smith, D. A., Brallier, S. A., & Palm, L. J. (2010). Computer-based testing: A comparison of computer-based and paper-and-pencil assessment. *Academy of Educational Leadership*, 14(4), 117-125.
- McDonald, A. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*, 39, 299-312.
- Messick, S. (1989). Validity. In R. Linn (Ed.) *Educational Measurement* (3rd ed.). London: Collier Macmillan Publishers.
- Mojarrad, H., Hemmati, F., JafariGohar, M., and Sadeghi, A. (2013). Computer-based assessment (CBA) Vs. paper/pencil-based assessment (PPBA): An investigation into the performance and attitude of Iranian EFL learners' Reading Comprehension. *International Journal of Language Learning and Applied Linguistic World*. 4(4),418-428.
- Noyes, J. M., Garland, K. J., & Robbins, L. (2004). Paper-based versus computer-based assessment: Is workload another test mode effect. *British Journal of Educational Technology*, 35(1), 111-13.
- Paek, P. (2005). Recent trends in comparability studies. *Pearson Educational Measurement Research Reports. Research Report 05-05. Pearson Educational Measurement. USA*.
- Pomplun, M., Frey, S., & Becker, D. (2002). The score equivalence of paper- and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62(2), 337- 354.
- Sam, H. K., Othman, A. E. A., & Nordin, Z. S. (2005). Computer self-efficacy, computer anxiety, and attitudes toward the internet: A study among undergraduates in UNIMAS. *Educational Technology & Society*, 8 (4), 205-219.
- Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*.5(2), 38-59
- Shohamy, E. (2001). *The Power of Tests – A Critical Perspective on the Uses of Language Tests*. Edinburgh: Pearson Education Ltd.
- Tanner, D. E. (2001). Authentic assessment: A solution, or part of the problem. *High School Journal*, 85(1), 6-13.
- Ward, T., Hooper, S., & Hannafin, K. (1989). The effects of computerized tests on the performance and attitudes of college students. *Journal of Educational Computing Research*, 5, 327-333.

Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006). *Score comparability of online and paper administrations of the Texas assessment of knowledge and skills*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Available at http://www.pearsonedmeasurement.com/downloads/research/RR_06_01.pdf.

Yurdabakan, I., & Uzunkavak, C. (2012). Primary school students' attitudes towards computer based testing and assessment in turkey. *Turkish Online Journal of Distance Education*, 13(3), 177-188.