

The investigation of Difference between PPT and CBT Results of EFL Learners in Iran: Computer Familiarity and Test Performance in CBT

Monirosadat Hosseini¹, Mohamad Jafre Zainol Abidin², Hamid Kamarzarrin³,
Mohamad Khaledian⁴

¹PhD student in University Sains Malaysia (USM), 11800, Penang, Malaysia

²Professor assistant at University Sains Malaysia (USM), 11800 Penang, Malaysia

³Assistant Professor, Department of Psychology, Faculty of Humanities, Payame Noor University,
PO BOX 19395 - 3697, Tehran, I.R. of Iran

⁴Department of Psychology, Faculty of Humanities, Payame Noor University, PO BOX 19395 - 3697,
Tehran, I.R. of Iran

E-mail address: Mohamad_khaledian22@yahoo.com

ABSTRACT

The purpose of this study is to examine the score comparability of institutional English reading tests in two testing methods, i.e. paper-based and computer-based tests taken by Iranian EFL learners in four language institutes and their branches in Iran. In the present study, the researcher tried to examine whether there is any difference between computer-based test results (henceforth CBT) and paper-based test (PPT) results of a reading comprehension test as well as exploring the relationship between students' prior computer experience and their test performance in CBT. Two equivalent tests were administered to one group of EFL learners in two different occasions, one in paper-based format and the other in computer-based test. Utilizing t-test, the means of two modes have been compared and the results showed the priority of PPT over CBT with .01 degree of difference at $p < 05$. Using ANOVA, the findings revealed that computer experience had no significant influence on the students' performance in computerized test.

Keywords: Computer-Based Test (CBT); Paper-and-Pencil Test (PBT); computer familiarity; test performance

1. INTRODUCTION

Technology is increasingly being promulgated as a powerful mechanism that can transform education. Technology is not a novel topic for language testers. With the appearance of new technologies, computerized testing has begun to be widespread and implemented in large scale testing (Higgins, et al., 2005). However, the limited accessibility to computer and high cost in the past, limited the implementation of computerized language testing. But nowadays, the accessibility to computers and widespread use of computers in educational settings makes computerized testing more versatile than before. Moreover, in language learning, the most precise and available way is through computers and on-line

process (Fleming & Hiple, 2004). Such developments in computer technologies have influenced many areas (Pommerich, 2004) one of which is learning and testing English through computer and sometimes Internet.

This is why some International testing organization conduct their examinations (such as TOEFL, IELTS, and the like) through computer offline or online. However, these exams are developed and administered internationally by English native language countries. However, in non-native countries such as Iran, where English is not first or even second language, developing and administering such tests is not popular and even thinking about it is somehow avoided because of the unfamiliarity with such exams. However, making students and English learners familiar with such exams before encountering with real computer based situations internationally (which is inevitable) is very necessary for English teachers and test developers, especially for English language institutes in a competitive era. It has been observed that some students after taking such computerized tests complain that their test score is not real representative of their language proficiency because of their unfamiliarity with such test modes.

However, as institutions started to accomplish computer-based testing in their examination systems along with traditionally paper-based testing systems, concerns arise about the comparability of scores from the two administration modes (Wang, 2004). As the computerized tests have been using for almost 20 years (Laborda, 2007), and the computer assisted language learning (CALL) has been common since the middle of 20th century, it has been necessary to develop the means to include computerized tests (Leahy, et al., 2005). Although as Zhang & Lau (2006) suggest that CBT offers many advantages over traditional PPT, assessment experts, researchers, practitioners, and educators have concerns about the equivalency of scores between the two test administration modes.

While computers have been important in language testing, only a relatively small group of professional language testers uses computers in producing and validating language tests. However, scores derived from CBT as compared to PPT might reflect not only the examinee's proficiency in the construct being measured, but also the level of language proficiency (Kathleen, 2006).

2. REVIEW OF RELATED STUDIES

Isleem (2003), in a research conducted on Ohioan technology education teachers, found that computer competence and experience were the strongest predictors of attitudes towards computer use and taking CBT. Likely, Berner (2003) has done a study on the importance of computer competence in determining teachers' attitudes towards ICT. The results of his study confirmed that computer competency was the most significant predictor of teachers' interest in using computer in education.

Many studies have investigated the effects of computer ownerships on the teachers' computer competence and concentration on improving computer attitudes and usage. Briefly, the results consistently correlate with attitudes towards using computers in examinations and positive effects for preparing teaching and learning materials (Roussos, 2007; Sadik, 2006). In specific cases, substantial differences between paper and computer-based testing may occur depending on the specific measure, the participants, and the soft and hardware realizations and in whole computer familiarity and competency (Bridgeman, Lennon, and Jackenthal 2003). They compared students' reading scores in a computer-based test as a function of different screen sizes and resolutions.

They found that small screens at low screen resolution impair reading performance and reasoned that scrolling caused the differences in performance. However, they suggest that the computer using experience could decrease the influence of such factors. Pomplun, Frey, & Becker (2002) in their study found that the response procedures, and not the characteristics of the presentation (e.g., screen resolution), are decisive for differences in reading performance across media. For example, clicking the correct answer with a mouse is more time-consuming than ticking the solution on a sheet with a pen, especially with speeded measures, this extra time is a disadvantage for participants completing a computerized test version.

This is related to computer competency and experience to be fast or slow unless scores are corrected for speediness. Rezaee and his colleagues (2012) have done an investigation on the relationship between attitude and computer using in teaching contexts in Malaysia. They found the relationship between computer competency and attitude towards using computer in educational contexts, either in teaching or in testing. Similarly, Horkay, Bennett, Allen, & Kaplan, (2005) have a study on the writing performance in a comparability study of CBT vs. PPT. In their study, Computer familiarity was assessed as (a) hands-on computer proficiency, (b) extent of computer use in general, and (c) computer use for writing in particular was significantly related to computer-based writing performance after controlling for the paper-based performance.

The results of numerous studies in the comparability of PPT and CBT show that there is no empirical evidence that identical paper-based and computer-based tests obtain the same results. The factors that may influence the test results instead of the construct being measured are referred to as the “test mode effect” (Clariana & Wallace, 2002). It means the factors that influence the test performance of examinees such as computer familiarity, attitude towards computer using, age, gender, and environmental contexts. These factors could be the results of the variety in the results of comparability studies of PPT and CBT.

For example, paper-based test scores were greater than computer-based test scores for both mathematics and English tests in Mazzeo et al.'s (1991) study. While computer-based test scores were greater than paper-based test scores for a dental hygiene course unit midterm examination (DeAngelis, 2000); and some studies, in contrast, have reported non-significant difference between computer and paper-based tests (Schaeffer et al., 1993; Mason, et al., 2001). In regard to such different results in comparability studies of PPT and CBT, Yurdabakan (2012) believes that even though computer accession opportunities increase students' computer competencies and CBT achievements (Bennett, et al. 2008), it is possible to evaluate that such approaches could be the reason of students' limited accession opportunities. Leeson (2006) identifies the factors lead to difficulties in CBT applications under two titles, as factors originating from “users” and “technology used”.

He states that the user's gender, the ability to process information, the ability to use a computer, and the level of anxiety could have an influence on an application. He gave the size and resolution of monitors, writing character and its length, the way the problem is presented, and having the option of review or not as technology originated factors. Many researchers have already done studies investigating the relationship between computer usage ability and achievement tests. Yurdabakan (2012) identifies some of these studies such as DeBell & Chapman (2003), Pomplun & Custer (2005), Pomplun et al (2006), and Bennett et al (2008), stressing that computer usage ability is an important predictor of respondent achievement; therefore, those poor students at computers may show low achievement in CBT. However, they believed that with the increase in computer technologies and access opportunities, such problems might decrease.

Boo (1997) in his study on the comparability of PPT and CBT suggested that there was no relationship between computer familiarity and test performance in three computerized reading tests. Taylor et al. (1999) also after examining the relationship between computer familiarity and test performance of 1,169 participants from different countries on TOEFL CBT, found no relationship between computer familiarity and examinees' test performance on TOEFL CBT. While increasing use of computer in academic contexts, especially in language learning, there have been many investigation on the comparability of test scores in two different test modes, some considering different key factors influencing test results such as computer familiarity, prior attitude towards using computers, age, gender, and some other factors.

The present study attempts to find out whether there is any relationship between computer experience and test performance in CBT in comparison with PPT among EFL learners in Iran. So the following research question has been raised and the researchers attempted to answer it.

RQ: Is there any relationship between Iranian EFL students' computer familiarity and their test performance in the computer-based test in comparison with paper-based equivalent tests? The results are in line with Boo (1997) and Taylor et al's (1999) in that there is no relationship between computer familiarity and test performance in CBT. The findings, on the other hand, is in contrast with DeBell & Chapman (2003), Pomplun & Custer (2005), Pomplun et al (2006), and Bennett et al (2008) who argue that there is strong relationship between computer familiarity and test performance in CBT. The results also confirm the difference between PPT and CBT results in that students perform better on paper-and-pencil based test.

3. PARTICIPANTS AND INSTRUMENTS

The participants were 162 EFL Iranian learners having been selected randomly from four English institutes and their branches in Tehran. Instructors agreed to cooperate and obtain the agreement of their students to participate in the study. All selected respondents were given a placement test to ensure their homogeneity in English proficiency to avoid any external variable influencing the test results.

Then after ensuring the homogeneity of students, they were given two equivalent multiple-choice tests derived from reading book to the participants in different occasions, one in paper and pencil type and the other in computerized format. In order to control the gender influence on the results, all participants were selected from males. In the next phase, a questionnaire on computer familiarity, validated by six English teachers, was distributed among participants to elicit their level of computer familiarity and experience.

The first part of the questionnaire elicited demographic information of participants including age and institute. The second section of the questionnaire consisted of six items to gauge the learners' familiarity. It was adopted from the Computer Familiarity Questionnaire (Eignor, Taylor, Kirsch, & Jamieson, 1998; Taylor et al., 1999) (e.g. "How often do you use a computer?") with Never, Once a Week or Less, More than Once a Week, and Do Not Know options. The questionnaire items were scored 1 for the Never response to 4 for More than Once a Week response, and Do Not Know responses receive no score.

Then, the score of CBT and the degree of learners' familiarity were compared to find out whether there is any relationship between computer familiarity and test results in computerized format. Pearson correlation was used to explore the strength of the relationship

between two variables (computer familiarity and test score), and t-test was applied to compare the mean score of two tests (PPT and CBT) to find out any difference if exists.

4. PROCEDURE

Two equivalent reading tests were given to one group in two different occasions, one in computer-based format, administered in laboratories equipped with adequate number of computers and intranet connections, and the other in paper-and-pencil version. The number of participants at the beginning was 162 but after collecting data, those completed data were used in the analysis, which turned to 106 real respondents.

The allotted time to answer the questions for both exams was 60 minutes. Before conducting the exam, the respondents were given some instruction about how to answer the computerized questions. It is worth mentioning that two equivalent tests had been piloted before administering the exams among similar respondents selected from the same institute. The results showed the high reliability between two tests (89 %). After conducting the exams, the participants were asked to fill out the questionnaire.

They were ensured that their responses would keep confidential and not be related to their test results and are just used to find out the relationship between their degree of familiarity and computerized test score.

5. RESULTS

To answer the research question, at first, descriptive statistics analysis was used to gain a better view of the data, and then the inferential statistics analysis was run to find out the relationship between mean scores. Table 1 shows the Statistical Tendency.

Table 1. Statistical Tendency of Mean Scores of PPT and CBT.

	Mean	N	Std. Deviation	Std. Error Mean
PPT	24.16	106	4.53376	.44036
CBT	23.16	106	5.85074	.56827

As displayed in Table 1, the students' mean score on the PPT, 24.16, is higher than their mean score on the CBT, 23.16. On the other hand, the standard deviation in PPT is lower than CBT. It means that the dispersion of the scores from mean score in CBT is higher than PPT; consequently, Standard Error of Measurement in PPT is lower than CBT. The results of the statistical tendency are shown in Figure 1.

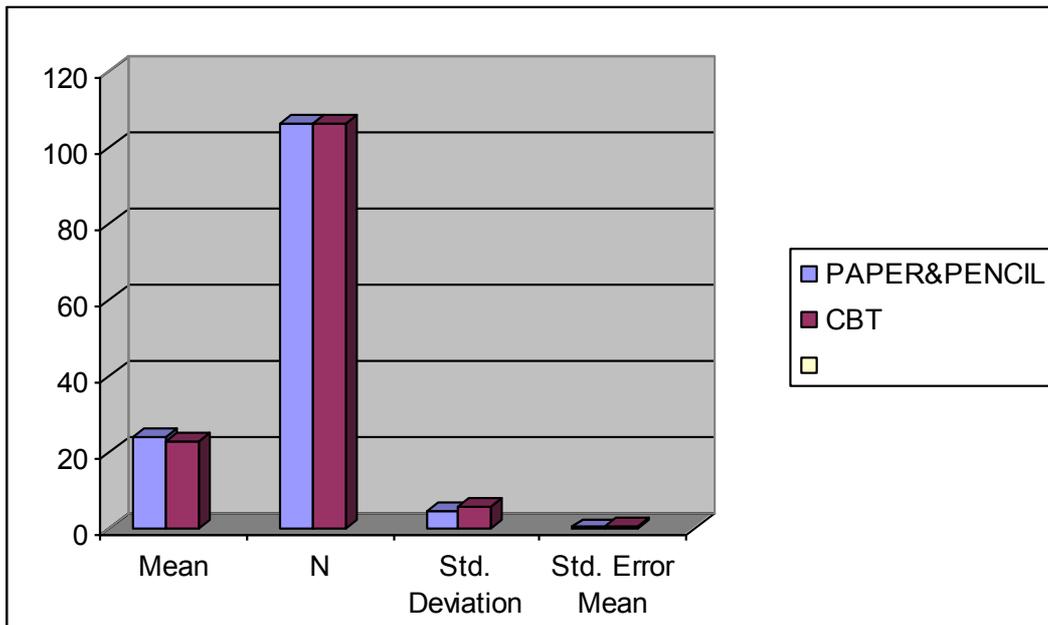


Figure 1. Bar graph showing the Statistical Tendency of two test methods.

Since the two means in the study came from the same subjects, paired sample t-test was run to compare the mean scores of the students on both tests. As can be seen from Table 2, the t-observed value is 1.99 at $P < 0.05$. This amount of t-value at 105 (N-1) degrees of freedom in $\alpha < .05$ is greater than the critical value of t, i.e. 1.98.

The results of inferential analysis showed difference between the students' mean scores on paper-based and computer-based tests, although it was not meaningful.

Table 2. Matched t-test results of PPT and CBT.

	T	D.F.	T-critical
PPT vs. CBT	1.991	105	1.98

In order to find out the relationship between computer familiarity and test performance of students in CBT, ANOVA has been run. As the results in Table 3 show, the F Observed value for the students' prior familiarity with computers and computer-based tests is 1.92 ($P = 0.14 > 0.05$). Based on these results, it can be concluded that the students' computer familiarity does not have any significant effect on students' test performance.

Previous results asserted that there is neither relationship nor interactive effect of prior computer familiarity of participants and their performance on computer-based tests. This implies that whether the subject has a high or low degree of computer familiarity, he/she would not be advantaged or disadvantaged when performing computer-based tests. Furthermore, this also supports the construct validity of the computer-based tests as this construct-irrelevant variable is not part of the construct measured by the computer-based tests.

Table 3. ANOVA results of interactive effect of computer familiarity on computer-based test scores.

Source		df	F	Sig.
mode * computer familiarity scale	Sphericity Assumed	1	1.92	.995

6. DISCUSSION AND CONCLUSIONS

The results of the study revealed that participants performed better on paper-based tests than computer-based test ($M_{PPT} = 24.6 > M_{CBT} = 23.6$). The findings is in contrast with those who argue that there is not any difference between CBT and PPT if the test administration condition is equivalent except the influence of students' preference in computer-based test rather than paper-based test mode (Bachman, 2000; Jamieson, 2005; Chapelle, 2007; Douglas & Hegelheimer, 2007). However, it supports the previous findings that favor the students' better performance in the paper-based test in comparison with computer-based tests (Coniam, 2006; Cumming, et al., 2006; Salimi et al. 2011). In addition, the results provide justification for including paper-based test administration. According the results of this study and other studies in confirming the priority of test results in paper-and-pencil test results, the whole problem may be the small capacity of the Working Memory (WM).

The whole brain is turned towards automaticity of mental processes. If the process is automatized, then few demands are made upon the WM and there is capacity available for higher mental processing. In other words, when bottom-up processing (here everything that goes below sometimes else such as eye contact, test environment, test format familiarity in PPT more than CBT, using keyboard or mouse instead of pen or pencil) is so hard and the brain is overloaded with it, little neutral networking is left for higher processes (such as focusing on the test content and thinking about the test responses).

To overcome such conditions, it is then necessary for all English institutions to habituate EFL students with computerized test contexts in order to make it automatized when performing CBT to minimize processing load of WM while taking English computerized tests. Therefore, the authors conclude that EFL students at least need to be at the trade-off level to be able to focus on content. As far as they are slaves of the forms, the content eludes from their focus, of course, almost all teachers are to some extent inclined or mentally forced to mingle lower and higher processing flaws in the scoring and this is one of the problems in doing research similar to ours. In short, how much automaticity is necessary for students to shift their attention from form to content is the high concern for the researchers in this area.

Moreover, computer technology has continued to be widespread into the 21st century as a crucial and versatile instrument for communication and education. However, rapid technological advancement can create a tendency towards acceptance of innovation and the belief that technology will be useful and solve all problems (Jamieson, 2005). This view can create problems, particularly if educators fail to act and react to the needs of learners.

References

- [1] Bachman L., *Language testing* 17(1) (2000) 1-42.
- [2] Bennett R. E., Braswell J., Oranje A., Sandene B., Kaplan B., Yan F., Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment* 6(9) (2008)
- [3] Berner, J. E. (2003). *A study of factors that may influence faculty in selected schools of education in the Commonwealth of Virginia to adopt computers in the classroom*. George Mason University.
- [4] Boo, J. (1997) *Computerized versus paper-and-pencil assessment of educational development: Score comparability and examinee preferences*. Unpublished PhD dissertation, University of Iowa, USA.
- [5] Bridgeman B., Lennon M. L., Jackenthal A. *Applied Measurement in Education* 16 (2003) 191-205.
- [6] Chapelle C. A., *Annual Review of Applied Linguistics* 27 (2007) 98-114.
- [7] Clariana R. Wallace P., *British Journal of Educational Technology* 33(5) (2002) 593-602.
- [8] Coniam, D. *ReCALL* 18 (2006) 193-211.
- [9] Cumming, A., R. Kantor, K. Baba, U. Erdosy & M. James (2006). Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for next generation TOEFL (TOEFL Monograph 30). Princeton, NJ: Educational Testing Service.
- [10] DeAngelis S., *Journal of Allied Health* 29 (2000) 161-164.
- [11] DeBell, M. & Chapman, C. (2003). *Computer and Internet use by children and adolescents in 2001: Statistical Analysis Report*. Washington, DC: National Center for Education Statistics.
- [12] Douglas D., Hegelheimer V., *Annual Review of Applied Linguistics* 27 (2007) 115-132.
- [13] Fleming, S. & Hipple, D. (2004). Foreign language distance education at the University of Hawai'i. In C. A. Spreen, (Ed.), *New technologies and language learning: issues and options* (Technical Report #25) (pp. 13-54). Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.
- [14] Higgins J., Russell M., & Hoffmann T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *Journal of Technology, Learning, and Assessment* 3(4) (2005), Retrieved from <http://escholarship.bc.edu/jtla/>
- [15] Horkay, N., Bennett, R. E., Allen, N., & Kaplan, B. (2005). Online assessment in writing. In B. Sandene, N. Horkay, R. E. Bennett, N. Allen, J. Braswell, B. Kaplan, & A. Oranje (Eds.), *Online assessment in mathematics and writing: Reports from the NAEP Technology-Based Assessment Project* (NCES 2005-457). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

-
- [16] Isleem M. (2003). *Relationships of selected factors and the level of computer use for instructional purposes by technology education teachers in Ohio public schools: a statewide survey*. The Ohio State University.
- [17] Jamieson J. M., *Annual Review of Applied Linguistics* 25 (2005) 228-242.
- [18] Kathleen Scalise and Bernard Gifford, (2006). *Computer-Based Assessment in E-Learning: A Framework for Constructing Intermediate Constraint Questions and Tasks for Technology Platforms* (Volume 4, Number 6).
- [19] Leahy S., Lyon C., Thompson M., William D., *Educational Leadership* 63(3) (2005) 19-24.
- [20] Laborda J. *Journal of LLT* 11 (2007) 3-9.
- [21] Leeson H., *International Journal of Testing* 6(1) (2006) 1-24.
- [22] Mason B. J., Patry M., Bernstein D. J., *Journal of Educational computing research* 24 (2001) 29-39.
- [23] Mazzeo, J. Druesne, B., Raffeld, P., Checketts, K. & Muhlstein, A. (1991). Comparability of computer and paper-and-pencil scores for two CLEP general examinations. (*College Board Report 91-5*). Princeton, NJ: ETS.
- [24] Pommerich M., Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *The Journal of Technology, Learning, and Assessment*, 2(6) (2004).
- [25] Pomplun M., Frey S., Becker D. F. *Educational and Psychological Measurement* 62 (2002) 337-354.
- [26] Pomplun M., Custer M., *Journal of Educational Computing Research* 32(2) (2005) 153-166.
- [27] Pomplun M., Ritchie T., Custer M., *Educational Assessment* 11(2) (2006) 127-143.
- [28] Rezaee A. A., Zainol Abidin M. J., Issa J. H., Mustafa P. O., *English Language Teaching* 5(1) (2012) 61-68.
- [29] Roussos P., *Computers in Human Behavior* 23(1) (2007) 578-590, <http://dx.doi.org/10.1016/j.chb.2004.10.027>.
- [30] Sadik A., *Education Review* 30(1) (2006) 86-113
- [31] Salimi, H., Rashidy, A., Salimi, A. H., Amini Farsani, M. (2011). Digitized and non-Digitized Language Assessment: A Comparative Study of Iranian EFL Language Learners. *International Conference on Languages, Literature and Linguistics IPEDR*, Vol. 26. IACSIT Press, Singapore.
- [32] Schaeffer, G., Reese, C., Steffen, M. McKinley, R. & Mills, C. (1993). Field Test of a Computer-Based GRE General Test. *Reports-Research/Technical* ETS-RR-93-07.
- [33] Taylor C., Kirsch I., Eignor D., Jamieson J., *Language Learning* 49(2) (1999) 219-274.

- [34] Wang, S. (2004). Online or paper: does delivery affect results? Administration mode comparability study for Stanford Diagnostic Reading and Mathematics Tests. Harcourt Assessment Inc, USA.
- [35] Yurdabakan I., *Turkish Online Journal of Distance Education* 13(3) (2012) 177-188
- [36] Zhang, L, & Lau, C. A. (2006, April). *A comparison study of testing mode using multiple-choice and constructed-response items – Lessons learned from a pilot study.* Paper presented at the Annual Meeting of the American Educational Association, San Francisco, CA.

(Received 03 September 2013; accepted 08 September 2013)