Evaluating the Comparability of PPT and CBT by Implementing the Compulsory Islamic Culture Course Test in Jordan University

Abdelnaser Sanad Alakyleh

Research Article

Evaluating the Comparability of PPT and CBT by Implementing the Compulsory Islamic Culture Course Test in Jordan University

Abdelnaser Sanad Alakyleh (iD)

Ministry of Education in Jordan/ Formerly Al-Jouf University

Abstract: Study aims to determine whether the university students' scores in the compulsory Islamic culture course test on a selected sample differ across the paper-and pencil test (PPT) & computer-based test (CBT) versions, and to reveal the relationship between gender and the student's level of performance in the test. Therefore, the study evaluated the comparability of two versions of a compulsory Islamic culture course test (PPTs) and (CBTs). The importance of conducting the study in Jordan stems from the fact that public and private universities have begun to move away from the traditional patterns of tests such (PPTs) and went towards (CBTs). In addition to detecting which model gives the best in the output and has the characteristics of the psychometric test, furthermore, indicates whether there were any differences between males and females, the study sample consisted of 120 individuals, 67 females and 53 males from scientific, health and humanities colleges. The results showed that there was no significant difference between the two versions provided to students CBT and PPT with 0.36 moderate correlation indicators in the pre-CBT test, no significant differences between the males and females in the CBT test results. Therefore, on the basis of the results of the present study, the CBT test is an option and a preferred alternative for regular students of the bachelor's level at the University of Jordan.

1. INTRODUCTION

CBT has recently appeared as one of the most demanded viable form of alternative assessment throughout the world. Along with the development of computer assisted language learning (CALL) in education, applying computers as accepted assessment tools seems to be inevitable especially in academic settings. In education, CBT is used to evaluate the language proficiency of English learners (Fleming & Hiple, 2004). Also, computer-based testing CBT has grown in popularity and will likely become the primary mode for delivering tests in the future. Computers revolutionized the world of training and development. Many investigators such as Fuhrer (1973) began researching on many points of mode which has enhanced training through computers. Many studies focused on the effects of using computers in the classroom for testing on various aspects of the learning environment such as student anxiety, teacher attitudes, student achievement and more.

*Corresponding Author E-mail: abd_275@yahoo.com

The computer had a significant impact on education over the past 20 years, its impact on educational testing is interesting and remarkable. Although a number of large educational institutions, such as ETS and English, Cambridge ESOL has designed CBTs, a limited number of educational institutions have adopted these tests. Few teachers apply them to their students, which explains the continued dominance of PPTs on the educational field. In the current period, the development of science and technology is advancing. This has an impact on life, including the education. The presence of technology in education is used to assist and improve the quality of learning (Woolfolk, 2007)., while the number of countries regard education as crucial for improving their current situation in every respect and moving it a step further in the information age of the 21st Century. In this context, Aslan (2006) pointed out that the developments which have occurred in information technology have given students fast and easy access to information, which has made a great contribution to education systems. As an example of the accelerated use of computers in the educational and academic fields mostly in tests, there have been several different versions of these examples and applications. That versions have become issues of interest to researchers and those interested educational, academic applications in the field of tests and comparisons with methods and traditional versions used by educators and academics to submit to examiners. With a view to carrying out the assessments of the examiners through its results on the applicable tests version in an effort to improve the quality and accuracy of subsequent decisions.

It is important to address two types of computer-based tests; Computer based standard testing CBTs and Computer-adaptive testing CATs. The CBT test is, in short, the usual paper version of the test, which has been converted into CBTs. Therefore CBT is as static as in the original paper copies of the test. In other words, all applicants for the computer test answer the questions in the same order in which questions are presented in the paper version, while a computer test adapted to the language proficiency of the CAT student, applicants answer different sets of questions, which are asked according to their level. Their answer affects a question about the following questions. A little bit of the first, and put it on the applicant to the test, and vice versa if the answer is wrong, the computer will choose an easier to difficult question, hence the name "adaptive test". CBTs are characterized by a number of features, tests are more stable and credible, and the CBT is superior to paper testing in many positive aspects. CAT has the ability to perform more rigorous and credible tests in determining the level of language knowledge among students. This is because it uses statistical analysis to assist the language test in identifying weak and good questions (Niemeyer, 1999)., but the problem with computerized tests arises when the matter of validity comes; however, there is no evidence to show that the construct of CBT may produce less valid tests. Instead, other factors may influence tests that have little to do with the testing objectives which the test developer intends to provide. For example, in many CBTs, it seems that the test designer started from a valid objective, but the limitations of the program, system, language or the tester's own characteristics have influenced the results of tests (Chapelle & Douglas, 2006).

Khoshsima, Hosseini, & Hashemi, (2017) explained that CBT has recently appeared as one of the most demanded viable form of alternative assessment throughout the world. Along with the development of computer assisted language learning (CALL) in education, applying computers as accepted assessment tools seem to be inevitable especially in academic settings, as mentioned (Holtzman, 1970) that IBM version 805 machine used in 1935 has been recorded as the first attempt to use computers in educational testing domain. It aimed to score objective multiple-choice item tests of American test takers each year to reduce the costs of scoring labor of millions of test takers throughout the USA, after publication of the first book on CBT in language domain. Al-amri, (2009) pointed out that many developments in technology caused rapid enhancements in comprehensive language testing software packages to use great advantages of CBT such as the innovation, efficiency and productivity, CBT assesses test

taker's language proficiency accurately by providing more efficient standardization of test administration conditions, in CBT the same instructions, materials and information are presented in an enhanced consistent and uniform way to all test takers, regardless of the testing population size, place and time of testing, but in some cases of large-scale CBT occasions, the security issues such as identity detection of test takers are the main concern.

Universities, some institutions, and testing organizations have started to change the mode of testing administration and to replace their paper and pencil tests PPTs with CBTs in language assessment field (Kate, 2012), while comparability and equivalency of test scores between the two test administration modes have been the real concerns for educators, scholars, practitioners and designers in assessment field (Lottridge, Nicewander, Schulz & Mitzel, 2008).

The sequence of studies and research on the preference of the examiners and educators to PPT compared to CBT, such as (Creed et al, 1987; Dillon, 1994; Clariana, 2005; Destefano, 2007; Dillon, 1994; Dundar, 2012 & Monirosadatet et al., 2014) study, showed that they agreed to prefer computer examiners CBT, while their results are better on paper and pen PPT, while (Higgins et al., 2005), (Al-amri, 2009) have been mentioned that there are no significant differences between the use of both models nor correlation between test mode preference and testing performance, used in the test and the performance of students, with regard to the gender of the respondent and his preference for any of the two models, some studies, such as (Gallagher, et al., 2002; Wallace & Clariana, 2005) indicated a preference for females to use the form PPT in front CBT model.

From the review of educational literature and previous studies that dealt with this important issue, the results of no significant differences between the use of both models, used in the test and the performance of students, shortage of correlation between test mode preference and testing performance, remains a subject of discussion and extensive examination of the different variables. Results of both versions affect variables such as gender, ethnic variables, motivation of the examiner, the concern of the test, the conditions of application, cognitive processes and technical issues which lead to the conclusion and the result that the use of the computer is not the tool of choice for evaluation.

Computers have become more widespread and used in academic aspects, especially in the application of tests in all its forms and their versions and in the results of which they depend on mainly the analysis of important decisions academically and practically, it has produced a lot of studies in the field of comparison between CBTs and paper and PPTs results are not compatible or consistent in the field of validity, reliability and significance differences of test scores.

Therefore, based on the above, the current study was to follow up and complete the research and study carried out by the researchers on the use of test models based on PPT compared to using CBT applying both models to a sample of university students and to a completely different topic of language. Focused on most studies in the application, and based on availability of data, potential and desire of volunteers from university students, the study came to discuss the comparison between the models of application on the subjects to confirm or deny or modify the previous studies of the results and analyzes.

2. METHODOLOGY

2.1. Purpose of The study

What may affect the validity of the effects of the test mode and the reliability of those results are not specific since the subjects of both male and female gender and their preference to test mode and performance will continue to discuss and research that the results of studies have varied between agreement and conflict on the subject and perhaps the proliferation of

computers in individual everyday life and make life more Automation. The increase in the use of computers in the academic community, especially in the field of tests, requires that traditional tests such as PPTs compared to CBTs waste time in preparation, processing, assessment and effort, as well as the tendency of the subjects often to computer tests. Equating the scores received from two types and suppressing test management, this may require further research on the relationship between some external variables of the mediator such as the sex test and test mode with test performance with greater attention, so the present study aims to determine whether the university students' scores in the compulsory Islamic culture course on a selected sample differ across the versions and to reveal the relationship between gender and the student's level of performance in the test, based on this purpose, the study derived the following questions:

> *RQ1*: Is there any statistically significant difference between PPTs and CBTs when applying of the Islamic culture course test for students of the University of Jordan?
>
> *RQ2*: Is there any significant difference in test results of CBT between female and male to Islamic culture course test on the students of the University of Jordan?
>
> *RQ3*: Do performance on CBT affected by participants' prior testing mode preferences?

2.2. Method

The present research that covered both comparison and correlational studies explored the comparability of paper and computer-based testing in a compulsory Islamic culture course and the correlation between some external moderator factors including test taker's characteristics such as computer attitude,. In order to reach solid conclusions in this research, a quantitative instrument's were used to investigate the difference between test results due to its advantages such as easy and fast data collection, consistency and accuracy of collected data and proper descriptive and inferential results, the study used the technique used by the (Khoshsima1, et al., 2017) study to examine the differences between the averages. The analysis of variance ANOVA was used in the study, with the different study population, sample size and nature of the test subject, and to reach the goals of the present study, a quantitative approach including descriptive statistics and was used to answer the first research question by comparing the means of sets of scores and to examine the significant difference between computer familiarity and attitudes, and testing performance of students, add to see if there was any difference between the scores of PPT and CBT. A majority of research conducted on PPT and CBT comparability study focused on the differences in means and standard deviations, (e.g. Makiney, Rosen, & Davis, 2003; Pinsoneault, 1996).

2.3. Population and sample study

The current study society consists of all the students of the University of Jordan for the academic year 2016/2017, which are 35359 students according to the Department of Admission and Registration at the University. The study sample consist of 120 students of both sexes from three faculties chosen by the simple random method with (67 females& 53 males) to ensure that the study community accurately represents the characteristics of the study community as well as and equal opportunities for the appearance of any student from the study community in the sample. The faculties of pharmacy, science and Sharia were selected from the health, scientific and human faculties respectively, according to the conditions of the test and the students' opinions to participate in the experiment until the final stages. As for the reason for selecting the number 120 for the size of the sample, the arithmetic average of one division was taken within the different faculties and there were 40 students. Therefore, for three selected colleges, 120 students were taken, the final number of the study sample. And how to invite these students to participate in the study has been the number of volunteers from colleges, the three who participated in the desire and fill their will and of both genders, male and female,

with the news of the nature of the study and its purpose and mechanism of procedure and applied conditions, the study sample agreed to participate in it.

2.4. Study instrument:

The current study used the final test of the Islamic culture course, which is a compulsory university requirement for all students. To compare the scores from both the CBT and PPT versions, PPT of the Islamic culture course was transferred to the computerized – based version that students will use when they sit for the final test. Another instrument to collect the research data concerning the third research question was a simple question mentioned at the bottom of test takers' exam paper and screen, i.e. would you prefer taking the test on: paper – no difference – computer.

2.5. Procedure:

The method of study begins in the first session of the final test. The students are given the PPT test form using the multiple-choice test format, which includes each item with five options: strongly agree, agree, neutral, disagree and strongly disagree. After the test, the students answered the question: Would you prefer taking the test on: paper-no difference–computer, this question may explore and illustrates the relationship between the preferred version of the test and the performance on the test, while the responses of the students examined were collected and scored. In order to eliminate overwork and stress from the effects of testing and the impact of experience and training and reduce it, the test was done on the computerized – based version after six weeks of testing PPT where the examiners explained oral and written instructions for students to test the computer version. The vast majority of Examine students have demonstrated understanding and prior knowledge with such instructions and how to respond to this type of testing. Each student was given 40 minutes to answer 60 items, with attention to not counting the time of oral and written instruction. The mechanism was to show only one item on the student's test screen. As with the PPT, the examiners have the option to return to any item for review and change the response in the computerized – based version test, the question of the third question was answered exactly as in the first phase of the test at the end of this test.

3. FINDINGS

After the testing and data collection and correction, statistical analysis was carried out using the statistical package for social sciences SPSS V: 22 was the first to verify the validity of the test submitted to the students through the experts validity. The test was presented to a group of specialists in the course content and measurement & evaluation specialists to make their observations on the test items, some of which were deleted or modified while the rest of the test items were kept by the Experts as they are, for the final test to remain in the 60 items. As for the reliability of the test, and because of the importance of the internal consistency of the study data collection instrument, the persistence of a Cronbach's α reliability method was calculated from the test results applied to the Examine students and the test versions, the results of the analysis were shown relatively high reliability coefficients (PPT, α=0.91 and CBT, α=88) (Table 1).

Table 1. Internal consistency reliability (Cronbach's coefficients of PPT & CBT)

Testing Mode	N of Questions	Cronbach's Alpha
PPT	50	0.91
CBT	50	0.88

The sample of the study was divided into 67 females and 53 males. In order to arrive at the answers to the current study questions, the analysis of the ANOVA was used by comparing means of sets of scores to reveal whether there were any differences between the grades of CBT and PPT. Perhaps the most important thing in the current study in the comparison is to find differences in means and standard deviations. With a relatively higher mean score for PPT than for CBT by 0.57 points (Table 2), also (Table 2) shows that the mean scores and standard deviations on the PPT version were (M=53.43, SD= 3.86), while they were relatively lower on the CBT version with (M = 50.12, SD = 3.06). We also note that the standard deviation of the PPT version is higher than that of the CBT version, which means the dispersion of scores from mean score in PPT was higher than in CBT, leading us to conclude that the Standard Error of Measurement (SEM) in the PPT version Above it in the CBT version, This means statistically that a more consistent version in its scores with less dispersion and standard deviation than a PPT version.

Statistical analyzes in (Table 4) showed that there are no significant differences in the scores between the two versions CBT& PPT at the level of statistical significance 0.01. Which supports the null hypothesis that there are no significant differences in the results of the Islamic culture course tests for the two versions CBT& PPT on the students of the University of Jordan.

Table 2. Descriptive Statistics

	N	Mean	S.D	S.E	99% C.I Interval for Mean	
					Lower Bound	Upper Bound
PPT	120	53.43	21.36	3.86	49.65	57.51
CBT	120	50.12	16.74	3.06	48.24	52.00
Total	240	51.78	19.05	3.21	50.76	52.80

The results of ANOVA analysis of the test sessions conducted on the subjects indicated that the significant value was 0.904 at $P > 0.01$. As this value reveals and illustrates disclosed no statistical significant differences between the scores of test groups resulting from the forms of the test in addition to that the scores of the respondents, also did not differ for the two versions at $P < 0.05$. Thus, the statistical analysis presented in (Table 2) shows that there are no statistically differences between the PPT version scores of the test (n= 120, M=53.43, SD= 3.86) and the scores of CBT version of the test (n = 120, M = 50.12, SD = 3.06), (Sig = 0.904, p> 0.01).

Table 3. Results of comparison of test scores received from PPT & CBT versions

	Sum of Square	D.F	Mean Square	F	Sig
Between Groups	5.824	1	5.824	0.013	0.904
Within Groups	16252.667		118	173.734	
Total	16266.154		119		

As for the question of the second study to show whether the scores of the CBT version for the female examiners differ from the results of the degrees of male examiners for the same version, in (Table 4) we note that the distribution of male and female test scores using the CBT version showed that the mean scores of male examiners have reached (M=52.43, SD= 28.36) which is relatively lower than the observed values of females who have reached (M=53.62, SD= 9.74), so the highest mean score was found in Female CBT, with a relatively higher mean score by More than one (1) point slightly. Conversely, the standard deviation of females was lower than that of males from the groups that provided the test CBT, which meant that the test

scores of females were higher than that of males on the CBT version; this raises the values of SEM of female test scores in CBT.

Table 4. Descriptive Statistics of male and female CBT test scores.

	N	Mean	S.D.	S.E.	99% C.I Interval for Mean	
					Lower Bound	Upper Bound
Male CBT	53	52.43	28.36	3.14	42.36	62.50
Female CBT	67	53.62	9.74	2.56	40.55	59.69
Total	120	51.28	26.94	2.23	39.86	62.70

As for the results of the analysis in (Table 5) of the scores of male and female examiners using the CBT version, it shows that the observed significant value was 0.884. This amount of the significant value at 119 (N-1) of degrees of freedom shows no significant differences between the two groups of scores at level 0.01. (Sig= 0.884, p>0.01), thus, one way ANOVA analysis showed that the differences between the male participants' scores in CBT version (n = 53, M = 52.43, SD = 28.36) and female participant scores in CBT version of the test (n = 67, M = 53.62, SD = 9.74) were not statistically significant. (Sig= .884, p>0.01).

Table 5. Results of comparing male and female CBT scores.

	Sum of Square	D.F	Mean Square	F	Sig
Between Groups	**6.224**	1	6.224	0.033	**0.884**
Withein Groups	**6355.224**	118	53.86		
Total	**6372.194**	119			

As for the preference of the test version and the performance of the test and to show the relationship between them, the study examined the Pearson product-moment correlation to reveal this relationship, the results shown in (Table 6) showed that there is moderate correlation of 0.36, which indicated the classification of (Evan, 1996), which means that the changes in pre- CBT preference were Moderately correlated with changes in examine scores on the CBT version. These results differ in terms of the existence of indicators of moderate correlation values with (Flowers et al., 2011, Higgins et al., 2005; & Khoshsima et al., 2017) results for the existence of weak indicators correlation values. This may be due to the difference in the subject of the test in that it has changed from language content to culture content as well as an increase in the sample size used by the current study in which the sample size was 30 individuals, of whom six (6) were female only in (H, Khoshsima et al., 2017) study as an example, but not limited to most of the studies reviewed by the literature of the current study.

Table 6. Pearson product-moment correlation of pre-CBT testing mode preference and mean of CBT scores

Pre-CBT testing mode	Pearson product-moment correlation	- 0.36
Preference	Sig (2-tailed)	0.502
	N	120

Correlation of pre-CBT testing mode preference and mean of CBT scores.

The study examined the Pearson product-moment correlation to reveal this relationship between post-CBT testing mode preference and CBT testing performance, the correlation results of the test group in (Table 7) showed no significant correlation, the correlation coefficient of Pearson observed from the analysis was weakly with amount of **-0.143**.

Table 7. Pearson product-moment correlation of post-CBT testing mode preference and mean of CBT scores.

Post-CBT testing mode	Pearson product-moment correlation	- 0.143
Preference	Sig (2- tailed)	0.462
	N	120

Correlation of post-CBT testing mode preference and mean of CBT scores.

Another step analysis of the results of the study was to examine whether the examiners have performed better performance of their preferred test versions depending on pre and post-CBT testing performance and its relationship to testing performance. The findings in (Table 8) showed that, those of CBT participants who preferred PPT version of the test (PPT performance, M=51.69) outperformed on CBT (M=66.11) and those who preferred CBT (PPT performance, M=50.18) performed better on PPT (M=59.41). While PPT participants who preferred PPT version of the test (PPT performance, M=50.32) in the PPT session outperformed on CBT (M=53.44) and those who preferred CBT version of the test (PPT performance, (M=51.63) performed better on PPT (CBT performance, M=47.76), and those who did not mind taking the test on either version, did better on CBT (M=54.46).

The findings showed that testing performance and testing mode preference of test takers had no positive interaction values, which means that testing mode preference inability to detect or influence the characteristics of the psychometric test, especially the validity of the test, the influence of exposure to the CBT version of the test on participants' posterior testing mode preference was examined.

Table 8. The relationship of pre-CBT testing mode preference of different preference groups with their testing performances

Testing sessions	Preferred testing mode	N	Mean		S. D	
			Pr-CBT p	Po-CBT p	Pre-CBT	Post-CBT
	Paper	75	50.32	53.44	16.74	28.20
PPTs	No difference	12	48.18	54.46	11.77	15.96
	Onscreen	33	51.63	47.76	26.89	17.94
	Paper	18	51.69	66.11	14.33	38.43
CBTs	No difference	14	46.87	52.88	15.45	15.66
	Onscreen	88	59.41	50.18	19.35	19.35

Note: Pr-CBT p refer to Pre-CBT performance and Po-CBT p refer to Post-CBT performance

To show the difference between testing mode preference before and after exposure to CBT, the answers of the participants to the testing mode preference question were collected to show proportion responses, (Table 9) values indicted that On-paper (Pre-CBT) PPT (n=75, P=625) while (Post-CBT) CBT (n=18, P= 15) however, no difference (Pre-CBT) PPT (n=12, P=10), while (Post-CBT) CBT (n=14, P=11.66), but for the On-screen (Pre-CBT) PPT (n= 33, P= 275), while (Post-CBT) CBT (n= 88, P= 73). Findings revealed that although test takers show high preference for taking CBT, they did better on PPT version of the test. We find that the number of participants who preferred to take PPT by reviewing these values from the results and those participants who preferred to take the test in either version changed for the side of the participants who preferred to take CBT.

Table 9. Differences between pre and post-CBT testing mode preferences

Preferred testing Mode	(Pre-CBT) PPT		(Post-CBT) CBT	
	Frequency	Percentage	Frequency	Percentage
On paper	75	62.5	18	15
No difference	12	10	14	11.66
Onscreen	33	27.5	88	73
Total	120	100	120	100

From (Table 9) we observe that: 62.5%, 27.5% of participants preferred to take PPT and CBT versions of the test, respectively, before the exposure to the CBT. Besides, 10% of participants didn't mind taking the test in either mode. After implementing CBT version of the test, only 15% still preferred to take PPT and 11.66% of the participants didn't mind taking the test in either mode. In this step of the study, the greatest percentage (73) was provided by the participants who chose CBT version of the test. The findings revealed that, after exposure to the CBT, the number of participants who preferred to take PPT and those participants who preferred to take the test in either mode changed in favor of the participants who preferred to take CBT.

4. CONCLUSION:

The present study was conducted for the purpose of investigating and determining whether there were any statistically significant differences in the scores of subjects obtained from the application of the compulsory islamic culture course test on the students of the University of Jordan and on the CBT and PPT versions. The results of the statistical analysis of the differences between females and males in performance on the test of the CBT version, indicated that there were no significant differences between the sexes in relation to there scores through the Two versions in the current study, it was found that sex differences were not a factor with a clear and strong performance on the subjects of both sexes effect.

This outcome is inconsistent with the findings of some studies of the no correlation indicator or a low correlation indicator either on the pre-CBT or post-CBT studies Such as (Flowers et al., 2011), (Higgins et al., 2005) and (Khoshsima et al., 2017). It is clear from the results of the present study, although the test takers CBT version may change its preference for the pre-test version, which may lead to acceptable performance relative to the type of test version, preferring the type of pre-test version as a moderate variable does not have that strong or influential effect on the examiners performance of the CBT version. The present study recommends further research and studies on the same subject taking into account the specialty of the examine, test anxiety, the number of test items, the test time implementing, and the cultural background of the examine, further replications of the study with more participants who are less homogeneous would be desirable thereafter. Conduct further studies to see if the tests give similar grades when administered in PPT or CBT forms. Furthermore, by examining item-level performance in addition to the performance of the test level, this study provided an opportunity to review differences in form at the item level.

Declaration:

Ethics approval and consent to participate: The study was conducted in accordance with the Declaration of Helsinki, and all participants provided written, informed consent to participate.

Consent for publication: There are no details on individuals reported within the manuscript.

Competing interests: The author declares that he has no competing interests.

Funding: The research was funded personally by the researcher

Availability of data and materials: Data and materials described in the manuscript are available by contacting the author of the article at (abd_275@yahoo.com).

5. REFERENCES:

Ackerman, R, & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology*, 17(1), 18-32.

Al-Amri, S. (2009). *Computer-based testing vs. paper-based testing: establishing the comparability of reading tests through the evolution of a new comparability version in a Saudi EFL context*. (Unpublished doctoral dissertation), University of Essex, England.

Anne, A., Walgermo, B., & Bronnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research,* 58 (2), 61-68.

Aslan, O. (2006). New way of learning: E-Learning, Fırat University, *Journal of Social Science*, 16(2), 121-131.

Chapelle, C. A, & Douglas, D. (2006). *Assessing language through computer technology*. New York: Cambridge University Press.

Clariana, R, & Wallace, P. (2005). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology,* 33 (5), 593–602.

Gallagher, A., Bridgeman, B., & Cahalan, C. (2002). The effect of computer-based tests on racial/ethnic and gender groups. *Journal of Educational Measurement*, 39 (1), 133-147.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behaviour & Information Technology*, 6 (1), 3–13.

Destefano, D, & Lefevre, J. (2007). Cognitive load in hypertext reading: *A review. Computers in Human Behaviour*, 23(3), 1616–1641.

Dillon, A. (1994). Designing usable electronic text: *Ergonomic aspects of human information usage*. London: Taylor & Francis.

Dundar, H, & Akcayır, M. (2012). Tablet vs. Paper: The effect on learners' reading erformance. *International Electronic Journal of Elementary Education,* 4 (3), 441-450.

Ekrem, S. (2014). Computer versus Paper-Based Reading: A Case Study in English Language Teaching Context, *Mevlana International Journal of Education,* 4 (1).

Evans, J. D. (1996). Straightforward statistics for the behavioral sciences. *Pacific Grove*, CA: Brooks/Cole Publishing.

Fuhrer, S. (1973). A comparison of a computer-assisted testing procedure and standardized testing as predictors of success in community college technical mathematics (Doctoral dissertation)*, New York University, (1973)*. Dissertation Abstracts International, 34(6), 3086.

Fleming, S., Hiple, D., & Du, Y. (2002). Foreign language distance education at the University of Hawaii. In C. A. Spreen (ed.), *New technologies and language learning: issues and options (Technical Report #25)* Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center, 13-54.

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.

Gallagher, A., Bridgeman, B., & Cahalan, C. (2002). The effect of computer-based tests on racial-ethnic and gender groups M, *Journal of Educational Measurement*, 39 (2), 133-147.

Khoshsima, M. H, Hosseini, S., & Hashemi, A. (2017). Cross-Mode Comparability of Computer-Based Testing (CBT) Versus Paper-Pencil Based Testing (PPT): An Investigation of Testing Administration Mode among Iranian Intermediate EFL LearnersToroujeni1, *English Language Teaching*, 10 (2), 64-72.

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), 1-36.

Holtzman, W. H. (1970). *Individually tailored testing: Discussion. In W. H. Holtzman (Ed.)*, Computer-assisted instruction, testing, and guidance. New York: Harper & Row.

Kate Tzu, C. C. (2012). Elementary EFL teachers' computer phobia and computer self-efficacy in Taiwan. TOJET: *The Turkish Online Journal of Educational Technology*, 11 (2), 100-107.

Kim, J. (2013). Reading from an LCD monitor versus paper: Teenagers' reading performance. *International Journal of Research Studies in Educational Technology*, (2) 1, 15-24.

Lottridge, S.M., Nicewander, W.A., Schulz, E.M., & Mitzel, H.C. (2008). *Comparability of paper-based and computer-based Tests*: A review of the methodology. Monterey, CA: Pacific Metrics Corporation, Section 2: Studies of Comparability Methods, 13-32 .

Makiney, J. D., Rosen, C., Davis, B. W., Tinios, K., & Young, P. (2003). Examining the measurement equivalence of paper and computerized job analyses scales. *18th Annual Conference of the Society for Industrial and Organizational Psychology*, Orlando, FL.

Mojarrad, H., Hemmati, F., Jafari Gohar, M.,& 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 Linguistics World*, 4 (4), 418-428.

Niemeyer, C. (1999). A Computerized Final Exam for a Library Skills Course. *Reference Services Review*, 27 (1), 90-106.

Pinsoneault, T. B. (1996). Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Inventory-2. Computers in Human Behavior, 12, 291-300.

Poggio, J., Glasnapp, D., Yang, X.,& Poggio, A. (2005). A Comparative Evaluation of Score Results from Computerised and Paper & Pencil Mathematics Testing in a Large Scale State Assessment Program, *The Journal of Technology Learning and Assessment*, 3 (6), 1-31.

Pommerich, M. (2004).Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *The Journal of Technology Learning, and Assessment*, 2(6), 1-45.

Salimi, H., Rashidy, A., Salimi, H.,& Amini, 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.

Woolfolk, A. (2007). *Educational psychology* (10th ed.).New York: Pearson Education, Inc.

Zandvliet, D.,& Farragher, P. (1997). A comparison of computer-administered and written tests, *Journal of Research on Computing in Education,* 29(4), 423-438.