Perceptions of Iranian EFL Learners towards Learning Language through Computers: An Investigation on the TOEFL (Paper and Computer Based) Test

Adnan Satariyan [1*]

[1] Faculty of Persian Literature and Foreign Languages, Islamic Azad University, South Tehran Branch, Tehran, Iran/ Faculty of Education, University of Tasmania (UTAS), Australia

Abstract

Computer technology has influenced the realm of language teaching and testing so drastically that no language teaching program could be imagined nowadays without the use digitized software and multimedia. This study aims at investigating Iranian EFL learners' performance on paper-based test compared with their performance on computer-based test while considering their attitudes towards the computer and learning language through computer. The sample selected for this study consisted of 205 Iranian male and female EFL learners, having been selected randomly from some language institutes and colleges, their age ranging from 17 to 27 years. To materialize the objectives of the study, the researchers used three research instruments: a test in two versions, a questionnaire and an interview. The Objective Placement Test (one computer-based and the other paper-based) including, listening, reading and language use was given to participants in two separate administrations. The format of the computer-based version of the test was designed by the researchers so that it could be the same as that of the paper-based version and could have the same level of practicality. The questionnaire was based on Min (1998), designed to measure the participants' attitudes towards the computer in general and computer-based language learning in particular. To confirm the questionnaire data, an interview was also randomly conducted with 20 learners. It was found that Iranian EFL learners are mostly exposed to paper-based tests. The findings revealed that although learners showed positive attitude towards computer-based tests and digitized language learning, they performed better on the paper-based test than on the computer-based test.

Keywords: EFL learners' attitudes, Computer-based language learning, Computer-based English test, Paper-based English test

INTRODUCTION

Language testing practice was evolved in the 1960s and 70s, essentially by a theoretical view of language ability as consisting of skills (speaking, listening, reading, and writing), and components (grammar, vocabulary, and pronunciation), and as an approach to test design. Also, language testing research was dominated by the hypothesis

*Corresponding Author's Email:
Adnan.Satariyan@utas.edu.au

that language proficiency consisted of a single unitary trait and a quantitative, statistical research methodology (Oller, 1979). The 1980s experienced other areas of development in language testing; throughout this period, second language acquisition research inclined language testers to uncover a variety of factors such as field independence/dependence (Chapelle, 1988), academic discipline and background knowledge (Hale, 1988) and discourse domains (Douglas and Selinker, 1985) on language test performance. In the 1990s, on the other hand, the field of language testing witnessed expansion in a number of areas: a) research methodology (criterion-referenced measurement, structural equation modeling, qualitative research approaches, generalizability theory, and item response theory), b) practical advances (testing cross-cultural pragmatics, testing languages for specific purposes, testing vocabulary, and computer-based assessment), c) research into factors that affect performance on language tests (characteristics of the testing procedure, the test-taking process, and characteristics of test-takers), d) performance assessment, and e) ethical issues (ethics of test use and professionalization of the field) (Bachman, 2007).

Inspired by the developments stated above, computer-based testing has been increasingly applied around the world (Chapelle, 2007; Bachman, 2000). Developments in language testing research in the past twenty years have witnessed progress in using computer technology in education and in language assessment including developing, scoring, administration, storing, handling test data, and performing sophisticated statistical analysis. The widespread use of computer technology in the delivery of language tests and the availability of personal computers, along with increased computer familiarity have made the administration of computer-based tests feasible for the first time on a large scale (Douglas, 2007). By the same token, advances in multimedia and web technology offer the potential for designing and developing computer-based tests that are more authentic and

interactional than their paper- and- pencil counterparts. For instance, in the computer-based TOEFL test administrations, a test taker would receive test items on the computer screen or based on a set order of items (as in a paper-and-pencil test). In both of these ways, the test-taker would need to have the computer literacy such as clicking, scrolling, highlighting, etc., in order to be able to read the test items and record the answers. Hence, the issue of the administration of computer test is under question if the test is required in areas where such computer skills and computer-based tests are relatively new or non-existent as they are in the context of the present study. Furthermore, other relevant issues need to be taken into account. For example, cost is a concern when the paper-and-pencil version of the test is replaced with computer-based version. Geographical consideration, along with the comparability of computer-based and paper- and-pencil tests, is a matter of concern as well. Thus, if these are the major concerns, then technology with computer-based testing may raise more questions than answers (Chapelle and Douglas, 2006; Douglas, 2000).

However, there remains a split among the researchers in the field as to whether test takers perform better on the computer-based tests or prefer them over the paper-and-pencil ones (Sawaki, 1999; Bachman, 2000; Jamieson, 2005). On the other hand, a number of studies have suggested that there is little difference between learners' performance on paper-and-pencil language tests (PBTs) and computer-based tests (CBTs) by groups of test takers (Breland, Lee, & Muraki, 2004; Coniam, 2006; Wolfe & manalo,2005). Apart from these, the researchers have not been able to find a single study directly investigating the effect of computer literacy on CBT performance, and the upshot is that we still do not know with any certainty how computer technology in language tests affects individual test takers performance.

Perhaps the most ubiquitous concern raised about technology for language assessment is that examinees' performance on a CALT may fail to

reflect the same construct as what other forms of assessment would measure. The potential problem addresses the inferences that can be made about examinees' ability in terms of their test performance. Of course, if a computer-based test yields results which are significantly different from the results of a parallel paper-and-pencil test, it is a threat only to the extent that score users intend the two scores to be equivalent. This unstated aim typically underlies the discussion of the potential threat, and therefore, the following is one way educational measurement specialists have expressed the problem: "If the fact that items are presented on a computer screen, rather than on a piece of paper, changes the mental processes required to respond correctly to the item, the validity of the inferences based on these scores may be changed" (Wainer, Dorans, Eignor, Flaugher, Green, Mislervy, Steinberg & Thissen, 2000, p. 16). As a case in point, Canale (1986) stated that computer use held out the prospect of providing a better means for measuring different language constructs than that which was attainable through traditional test methods. However, research and development has had the intention of focusing on the goals of increasing efficiency and authenticity of testing, whereas to date few researchers have explored the intriguing question of how the computer might be used to assess different abilities, or constructs, than those currently assessed by traditional methods.

According to Douglas and Hegelheimer (2007), the shift from paper-based to computer-based tests must be considered to attain a better measure of the construct, not simply a more efficient one, which brings us to a consideration of the potential of computers to provide what Jamieson (2005) called "computerized tasks that better represent authentic language use" (p. 233). Furthermore, Chapelle and Douglas (2006) have suggested that "communicative language ability needs to be conceived in view of the joint role that language and technology play in the process of communication" (p.108), and recommended that language ability needs to be defined in terms of an interface between language and technology: "the ability to select and deploy appropriate language through the technology that are germane to a situation" (p.107).

Controversies on PBTs vs. CBTs

Perhaps the most tangible way of investigating the issue of whether examinees perform well on a computer-based test for the wrong reason (i.e., differential test performance due to factors other than differences in the ability to be measured) is through a study that compares examinees' performance on two tests which are the same except for the mode of delivery, i.e., one form of the test is delivered as a paper-and-pencil test and the other is administered by the computer. Amongst the first large-scale testing programs in the United States to transform their tests to computer-based testing, the Graduate Record Examination (GRE) carried out a number of comparisons on test items, sections, and total test scores in a research program aimed to investigate the comparability of the computer-based and paper-and-pencil forms of the GRE. In several studies which obtained test performance data from examinees who had taken both the computer-based form and the paper-and-pencil version of the GRE, researchers found very few and slight differences that were thought to warrant further investigation of mode effects, but they did not find general, obvious, and consistent mode effects that would suggest different inferences could be made from the two forms of the GRE (Schaeffer, Reese, Steffen, McKinley & Mills, 1993).

Chapelle and Douglas (2006) believed that learners' performance on a computer-based test might fail to reflect the same construct as what would be measured by other forms of assessment. With regard to this statement, studies comparing performance on CBT with that on the alternative delivery format, i.e., PBT (Choi et al., 2003, Coniam, 2006) have indicated rather small or mixed differences. Coniam (2006), for instance, explored some significant differences in learners' performance on a listening test administered by a computer compared with that on a paper-based

listening test resembling the CBT version. In a study conducted by Taylor, Jamieson, and Eignor (2000) local differences in computer use by the learners were reported. Developers of computer-based tests should be cautious about carefully investigating differences in computer familiarity among their prospective test takers. In a similar vein, Cumming et al (2006) worked out major discrepancies in the essays composed by integrated prompts compared to those produced in response to the traditional TOEFL essay. It might be inferred that their findings provide justification for including both task types on the iBT. Another study made a similar comparison applying a range of empirical methods that yielded complementary perspectives. In an effort to find out evidence, Choi, Kim, and Boo (2003) compared the paper-based language test with the computer version of the Test of English Proficiency developed by Seoul National University conducting content analysis, correlational analyses, ANOVA, and confirmatory factor analysis. Findings demonstrated considerable similarities between the two versions (PBT and CBT) of each of the different parts of the test, with the grammar sections showing the greatest similarities and the reading section displaying the largest differences.

It may be that if the profession is to appreciate the significance of detailed results of studies explaining differences between CALT (Computer Assisted Language Testing) and other forms of tests, statistical differences need to be found in test scores. As both Sawaki's and Chalhoub-Deville and Deville's (1999) reviews of research on CALT point out, despite the many CALT projects, no published research has attempted to investigate questions of score comparability. Chalhoub-Deville and Deville conclude that "research in L2 is still scarce regarding the comparability of PBT and CBT" (1999, p. 282). This conclusion may imply that comparability research is forthcoming, and indeed, some studies have appeared since then. However, considering that computer-based L2 testing has been going on for at least two decades, we have

to question why such research has not played a more prominent role in test developers' agendas. The aforementioned studies on the comparability of learners' performance on PBTs and CBTs on the one hand and the fact that the current study is the first large scale research on CALT in Iran on the other, the researchers were motivated to do this significant study in the context of Iran. Consequently, the researchers try to provide logical answers to the following research questions:

RQ1: Is there any statistically significant difference between Iranian learners' performance on a test administered by the computer and their performance on the traditional paper-based test?

RQ2: Which type of tests (paper-based tests or computer-based tests) is used more frequently by Iranian EFL learners?

RQ3: Do Iranian learners' attitudes towards computer, attitudes towards learning language through computers have any effect on their performance on the CBT test?

Method

Participants

The participants were 205 undergraduate university students as well as some English learners from different language institutes. The undergraduate university students were selected from University in Tehran. Instructors agreed to cooperate and get the consent of their students to partake in the study. The students were pursuing their bachelor's degree in English translation as this is one of the most popular undergraduate degrees offered by Iranian universities across the country. Nonetheless, given that the participants were not selected randomly from all of the English departments across the country, the researchers may not be able to make strong claims about the generalizability of the findings. The participants ranged in age from 17 to 27. It is noteworthy that a questionnaire and two test types-computer based and paper based tests were administered to 280 students, but 80 students were removed from the analysis because they had ei-

ther missing questionnaires, incomplete PBTs or CBTs.

Instrumentation

In order to measure the participants' attitudes towards computer-enhanced instructional program of the English, the researchers used Min's (1998) questionnaire. The first section of the questionnaire provided information about its purpose and elicited background knowledge on the participants' age, gender and educational level. The second part of the questionnaire consisted of 30 items to gauge the learners' attitudes towards learning English through computer. In effect, the questionnaire comprised of 30 items, 15 of which measured attitudes toward computer and the other 15 items measured participants' attitudes toward learning English through computer. What is more, the researchers distributed the 70-item Objective Placement Test developed by Jack C. Richards. The Objective Placement Test (both computer-based and paper-based test) has three sections: Listening, Reading, and Language Use. The participants were allowed 50 minutes to complete the test.

Procedures

In the beginning phase of the study, the Objective Placement Test (paper-based mode of assessment) was given to the participants. This test has three sections: In section I, the Listening section, nine conversations were played and participants were supposed to answer one or more questions about each one. The participants read, listened and answered the questions after the conversation ended. The participants were briefed about how to answer each section of the test on the answer sheet. They had 15 minutes to complete this section. Section II, the Reading section, has several short passages. After reading each passage, they were required to choose the correct answer for each question and mark it on the answer sheet.

They had 20 minutes to complete this section. Section III, the Language Use section, has 30 items. They had 15 minutes to complete this section. After a two-week time period, the second phase of the study took place. The participants were first asked to fill out the questionnaire of attitude towards computer and learning language through computer, and then they were given the Objective Placement Test (computer-based mode of assessment). The researchers in the present study designed the software of the Objective Placement Test which took one month to develop. Having piloted the Objective Placement Test by 10 experts, the researchers administered this version of test among the participants. To confirm the questionnaire findings, 20 participants were selected out of a total number of 205 learners who filled out the questionnaire and they agreed to be interviewed. The interviews, which were limited to 5-12 minutes to keep it manageable, were conducted in Persian individually after the questionnaire data were collected.

Analysis of the first research question

> *RQ1: Is there any statistically significant difference between Iranian learners' performance on test administered by computer and their performance on the traditional paper-based test?*

To probe the first question of the study, multivariate analysis of variance (MANOVA) was conducted. As can be seen from table1, the F Observed value comparing the students performance on paper- and computer- based tests is 6.40 (P = .012 < 0.05). The results of MANOVA showed a significance difference between the students mean scores on paper-based and computer-based tests.

Table 1
Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
Base	Pillai's Trace	.032	6.408[a]	1.000	196.000	.012
	Wilks' Lambda	.968	6.408[a]	1.000	196.000	.012
	Hotelling's Trace	.033	6.408[a]	1.000	196.000	.012
	Roy's Largest Root	.033	6.408[a]	1.000	196.000	.012
Base * AttCompLevel	Pillai's Trace	.002	.214[a]	2.000	196.000	.808
	Wilks' Lambda	.998	.214[a]	2.000	196.000	.808
	Hotelling's Trace	.002	.214[a]	2.000	196.000	.808
	Roy's Largest Root	.002	.214[a]	2.000	196.000	.808
Base * AttLearnLevel	Pillai's Trace	.006	.549[a]	2.000	196.000	.578
	Wilks' Lambda	.994	.549[a]	2.000	196.000	.578
	Hotelling's Trace	.006	.549[a]	2.000	196.000	.578
	Roy's Largest Root	.006	.549[a]	2.000	196.000	.578
Base * AttCompLevel * AttLearnLevel	Pillai's Trace	.002	.191[a]	2.000	196.000	.827
	Wilks' Lambda	.998	.191[a]	2.000	196.000	.827
	Hotelling's Trace	.002	.191[a]	2.000	196.000	.827
	Roy's Largest Root	.002	.191[a]	2.000	196.000	.827

a. Exact statistic

b. Design: Intercept + AttCompLevel + AttLearnLevel + AttCompLevel * AttLearnLevel

Within Subjects Design: Base

Analysis of the second research question

RQ2: Which type of tests (paper-based tests or computer-based tests) is used more frequently by Iranian EFL learners?

As displayed in Table 2, descriptive statistics showed that the Iranian EFL undergraduate, in the main, performed better on paper-based tests because the mean score of students who took the paper-based exceeds that of their performance on computer-based tests. (M= 41.54)

Table 2
Base: Descriptive statistics of learners` performance on tests
Measure: MEASURE_1

Base	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	41.540[a]	1.606	38.374	44.707
2	38.879[a]	1.530	35.860	41.897

a. Based on modified population marginal mean.

Analysis of the third research question

RQ3: Do Iranian learners' attitudes towards computer, attitudes towards learning language through computers have any effect on the performance of the students on the test?

As for the third question of the study, the results in Table 3 show that the F Observed value for the effect of the students' attitude toward computer on their performance on paper and computer based proficiency tests is 1.92 (P = .14 > 0.05). Based on these results it can be concluded that the students' attitudes toward computer does not have any significant effect on their performance on the proficiency test.

Table 3
Tests of Between-Subjects Effects
Measure: MEASURE_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	133125.947	1	133125.947	552.685	.000
At Comp Level	926.115	2	463.058	1.922	.149
At Learn Level	853.366	2	426.683	1.771	.173
At Comp Level * At Learn Level	558.917	2	279.459	1.160	.316
Error	47210.778	196	240.871		

As displayed in Table 4, the mean score for low, moderate, and high attitudes toward computer mean scores are 43.38, 36.65, and 42.37 respectively.

Table 4
AttCompLevel
Measure:MEASURE_1

At Comp Level	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Low	43.380[a]	3.313	36.846	49.914
Moderate	36.655	2.000	32.709	40.600
High	42.371[a]	2.697	37.051	47.691

a. Based on modified population marginal mean.

The F Observed value for the effect of the students' attitude towards learning through computer on their performance on paper and computer based proficiency tests is 1.77 (P = .17 > 0.05). Based on these results it can be concluded that there students attitude toward learning through computer does not have any significant effect on their performance on the proficiency test.

As displayed in Table 5, the mean score for low, moderate, and high attitude toward computer mean scores are 41.26, 42.71, and 35.40 respectively.

Table 5
AttLearnLevel
Measure:MEASURE_1

Att Learn Level	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Low	41.265[a]	3.574	34.217	48.314
Moderate	42.712	1.041	40.658	44.766
High	35.400[a]	3.470	28.556	42.244

a. Based on modified population marginal mean.

In sum, the following results were reported by the researchers:

There is not any significant interaction between;

a. Type of test and attitude toward computer (f = .21, p = .80 > .05)

b. Type of test and attitude toward learning through computer (F = .54, p = .57 > .05)

c. Attitude toward computer and Attitude to-

ward learning through computer ($F = 1.16$, $p = .31 > .05$)

d. Type of test, Attitude toward computer and Attitude toward learning through computer ($F = .19$, $p = .82 > .05$)

Results and Discussions

The results of the current study demonstrated significant difference between the Iranian EFL learners mean scores on paper-based test and computer-based counterpart. These learners performed better on paper-based tests (M=41.54).
These findings are in contrast with the research in the previous studies, which claim the preferences of the learners toward computer-based test administrationratherthantraditionalpaperbased. (Chapelle,2007; Bachman,2000; Douglas,2007; Jamieson,2005;Canale,1986;Douglas & Hegelheimer,2007).Nonetheless, it supports the literature in the sense that there is a significant difference in learners' performance on a computer-based test compared with that on a paper-based test favoring paper-based test. (Coniam, 2006; Cumminget al,2006). Therefore, these disparities in administration of the test provide support for the claim mentioned in the literature as there are disparities among the scholars in the field regarding whether test takers either perform better or preferred computer-based as opposed to traditional paper-based tests. (Sawaki,1999; Bachman,2000;Wolfe& Manalo,2005).Therefore, the researchers would expect to include both modes of presentation in the language program. However, the results provide justification for including paper-based test administration.

Moreover, computer technology has continued into the 21st century as a critical and powerful tool for communication. However, rapid technological advancement can create a tendency towards a blind acceptance of innovation and the belief that technology will solve all problems.(Kim,1997).This view can create obstacles, particularly if educators fail to act and react to the needs of learners. In Iran, it is perhaps becom-

ing more acceptable to learn English language using computers. Thus, students who do not hold positive attitudes towards learning through computers will be at a distinct disadvantage. Based on the results of the current study, the Iranian EFL learners showed positive attitudes towards learning through computers but they performed better on the paper- and -pencil based test administration.

The purpose of the current study was to perform an effective comparison of computerized assessment and a traditional based paper and pencil assessment in Iran. The study is unique in Iranian EFL context because it is the first that look into testing by computer juxtaposed traditional paper and pencil formats. There simply has not been research that compares these two groups in Iran. The findings of this study are in line with the claims mentioned by Canale (1986) and Laurier (1991) who pointed out that the challenge in applying technology to language assessment would be to figure out the benefits and limitations of the technologies in the context. One of the major limitations is test takers' apprehension about computer literacy and skills that might affect Iranian learners' performance. Lack of provision, cost, practicality, time, teachers' training and learners' briefing are among some limitations hinder the learners' preference for better performance on paper-based tests. Regarding the learners' computer literacy, Warschauer (1998) argues that computer literacy is of vital importance for success.

The results of the study showed the positive attitudes towards computerized language program. The learners reported the followings as the main reasons for their positive attitudes: ease of response, use of individually controlled time limits, and feedback. On the other hand, we have students who express anxiety as the main reason which impedes the learners' performance on computerized assessment effectively.

References

Bachman, L. F. Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing 2000, 17.1:* 1–42.

Breland, H., Lee, Y., & Muraki, E. Comparability of TOEFL CBT writing prompts: Response mode analyses (TOEFL Research Report No. RR-75). *Princeton, NJ: ETS,* 2004

Canale, M. The promise and threat of computerized adaptive assessment of reading comprehension. *In C.* Stansfield (Ed.), *Technology and language testing* (pp. 30–45). Washington, DC: TESOL Publications, 1986.

Chalhoub-Deville, M., & Deville, C. Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics. 1999, 19:* 273–299.

Chapelle, C. A. Technology and second language acquisition. *Annual Review of Applied Linguistics. 2007, 27:* 98-114.

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

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

Cumming, A., R. Kantor, K. Baba, U. Erdosy & M. James. *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, 2006.

Douglas, D. *Assessing language for specific purposes: theory and practice.* Cambridge: Cambridge University Press, 2000.

Douglas, D. & V. Hegelheimer. Assessing language using computer technology. *Annual Review of Applied Linguistics. 2007, 27:* 115–132.

Jamieson, J. [M.].Trends in computer-based second language assessment. *Annual Review of Applied Linguistics. 2005, 25:* 228–242.

Kim, Y. M. Language teaching enhanced by technology: Technopia or technicism? *Journal of the Applied Linguistics Association of Korea. 1997, 3:* 163-193.

Laurier, M. What we can do with computerized adaptive testing and what we cannot do! In Anivan, S., editor, *Current developments in language testing.* Singapore: *SEAMEO Regional Language Centre.*1991, 244–355.

Sawaki, Y. Reading in print and from computer screens: comparability of reading tests administered as paper-and-pencil and computer based tests. *Unpublished manuscript, Department of Applied Linguistics and TESL, University of California, Los Angeles,* 1999.

Warschauer, M. Researching technology in TESOL: Determinist, instrumental, and critical approaches. *TESOL Quarterly. 1998, 32(4):* 757–61.

Wolfe, E. W., & Manalo, J. R. *An investigation of the impact of composition medium on the quality of TOEFL writing scores. Princeton, NJ: ETS*, 2005.