A Comparative Study of IELTS Academic Writing Test in Paper Mode vs. on the Computer among EFL IELTS candidates at Tehran University[1]

Estudio comparativo de la prueba de redacción académica IELTS por escrito frente a la realizada en computadora en participantes de IELTS de inglés como lengua extranjera en la Universidad de Teherán

MEHDI DASTPAK

MOHAMMAD JAVAD RIASATI

MOHAMMAD SADEGH BAGHERI

EHSAN HADIPOUS

Jahrom University of Medical Sciences, Jahrom

Iran

md_dataforse@yahoo.com

Mjriasati2002@yahoo.com

bagheries@gmail.com

e.ehadipour@gmail.com

Resumen. El estudio actual es un intento para investigar si los estudiantes se desempeñaron de manera diferente en la prueba de escritura del Sistema Internacional de Evaluación del Idioma Inglés (IELTS) tanto por escrito como en computadora en términos de respuesta/logro de tareas, coherencia/cohesión, recurso léxico, rango gramatical y precisión. Además, explora si la familiaridad con la computadora de los candidatos era diferente en los grupos por escrito y en computadora. Para este propósito, se seleccionó un total de 108 candidatos de un total de 144 basándose en los resultados de la Prueba de Colocación de Oxford (OPT) en la Universidad de Teherán, Irán. Para recopilar los datos, se administró una muestra de redacción académica ya retirada del IELTS y un cuestionario de familiaridad con la computadora. Los participantes se dividieron en dos grupos iguales. En el grupo de Modo Escrito (PM), a los estudiantes se les dio la prueba para escribir en el modo de papel convencional. En el otro grupo Modo Computadora (CM), los estudiantes recibieron la misma prueba; sin embargo, se les pidió que escribieran la prueba en la computadora que se les proporcionó en su clase. Además, todos los participantes tomaron el cuestionario de familiaridad con la computadora. Los datos recopilados se analizaron mediante la prueba t de muestras independientes. Los hallazgos revelaron diferencias significativas entre el modo escrito y el modo en computadora en ambas tareas de escritura. Además, el análisis del cuestionario mostró el impacto de la familiaridad con la computadora de los candidatos en su desempeño en la escritura.

Palabras clave: *prueba de escritura de alto impacto; modo en papel; modo en computadora; familiaridad con la computadora.*

Abstract. The current study is an attempt to investigate whether learners perform differently on paper or on the computer in the International English Language Testing System (IELTS) writing test, in terms of Task response/achievement, coherence/cohesion, lexical resource, grammatical range and accuracy. In addition, it explores whether the candidates' computer familiarity are different in paper or computer groups. To this end, a total number of 108 candidates were selected out of 144 based on the results of the Oxford Placement Test (OPT) in Tehran University, Iran. To gather the data, a retired IELTS academic writing sample and a computer familiarity questionnaire were administered. The participants were divided into two equal groups. In the Paper Mode (PM) group, students were given the test to write conventionally on paper. In the other, Computer Mode (CM) group, the students were given the same test; but were asked to type the test in the computer provided for them in their class. Also, all the participants took the computer familiarity questionnaire. The gathered data were analyzed through the Independent samples t-test. The findings reveal significant differences between paper-based and computer-based modes in both writing tasks. Moreover, the analysis of the questionnaire shows the impact of the candidates' computer familiarity on their writing performance.

Keywords: *High-Stakes writing test; paper mode; computer mode; computer familiarity.*

2

Introduction

According to Davies (2007), IELTS as a high-stakes test, greatly influences the educational and professional life and success of candidates. The two modules of IELTS, namely General and Academic, are assumed for admission to academic centers and immigration issues respectively. As Quaid (2018) expresses, the test is branded as a general proficiency theoretical model which means that underlying the test is a belief that there is some indivisible body of language knowledge within each test taker that is technically analyzable. This knowledge includes a fixed proficiency construct that is present in participants in the assessment process. Based on this knowledge of language, each individual can be rated (Fulcher, 2014). The IELTS test sorts no functional or structural syllabus to model (Quaid, 2018), while it is assumed that the test takers' performance can be generalized to the real world.

The performance of each test-taker is simplified into an overall score and also four elemental scores, to be simply and efficiently interpreted by various stakeholders (O'Loughlin, 2011). In listening and reading, candidates obtain raw scores out of 40 which are equal to a 1 to 9 band score. However, the 1 to 9 band scores in Writing and Speaking are given by an examiner in four sub-criteria (IELTS, 2014), which is rounded to the nearest 0.5 band if required. In the IELTS test both of the tasks are scored by two different examiners, which leads to a total band score that comprises an average of the two, with task two weighing more as it is longer. Ultimately, the average of the four sub-tests, which are also called band descriptors, is awarded corresponding to a description of proficiency provided by IELTS (IELTS, 2014). The five scores (numbers) which candidates receive constitute the only performance information they obtain from the test.

Review of Literature

Historical Evolution of Testing English

It has been seventy-five years since the British Council has been involved in language testing. They have spread their influence worldwide through exporting English language exams and skills in language assessment. Seven years after being founded (1934), the Council moved into the world of language testing by signing a contract with the University of Cambridge Local Examination Syndicate (UCLES). They agreed that the British Council would provide EFL expertise and also financial and technical support in order to help UCLES develop their set of English tests. The high points of this phase were the British Council inspired Cambridge Diploma of English Studies, introduced in the 1940s, and the significant role of the British Council in the development of English Language Testing Service (ELTS) in the 1970s, which is the origin to the IELTS. Developing indigenous English language tests worldwide by the British Council during the last thirty years has boosted the advancement of English.

It was in the early 1990s that the efforts of the British Council switched to delivering British examinations through its network, which was accessible worldwide, from its previous test development. Though, the organization considered returning to test development by the early of the 21st century. Then, an in-house test was developed for the first time in thirty years. This set the stage for the emergence of professional expertise in the testing of language by the British Counsel, and strategic influence grew for the organization on English language assessment. This influence is based on a commitment to accessible and affordable tests, and also the efficient delivery of tests whose reporting and marking was strengthened   by innovative approaches to language testing. The return, which is so clear by British Counsel, can be considered as a tool to enrich Britain's soft power.

Writing mode and writing performance

A number of researchers have found that language testing has gained more significance, and it has turned out to be stronger in recent years. Language testing and its results play a vital role in candidates' lives (Uysal, 2010). The results of language tests can affect candidates' social and academic or educational life (Shohamy, 2001a; Puspawati, 2012). It is the test result that determines students' learning when it comes to their educational life. It can be mentioned that their education is based on tests. Tests come from the materials such as books that are provided to students, and the course of learning is aimed at tests. Shohamy (2001a) stresses that "tests determine what candidates need to know, what they need to learn, and what they have learned"(p.17).

The application of computer-based education and assessment is receiving more attention (Poggio, 2005; Lottridge. 2008; Yurdabakan, 2012). However, the results of numerous studies in comparing paper mode (PM) and computer mode (CM) have resulted in no empirical evidence that CM and PM tests lead to the same results. Clariana and Wallace (2002) mention that some factors, other than the construct being measured, may affect the results of such tests.

Some studies have investigated the performance of individuals by comparing their scores and/or writing processes in both computer and paper modes (e.g., Barkaoui, 2016; Blackhurst, 2005; Breland, Lee, &Muraki, 2004; Green &Maycock, 2004; Horkay, Bennett, Allen, Kaplan, & Yan, 2006; Jin & Yan, 2017; Lee, 2002; Li, 2006; Russell & Haney, 1997; Weir, O'Sullivan, and Jin (2007); Wolfe & Manalo, 2005). The matter that these lines of research address is to determine whether the scores on CM assessment show the same ability on the part of the participants as those in PM tests, which are supposed to be equal (Chapelle& Douglas, 2006). To exemplify, three independent investigations made a comparative study on the differences between scores in computer mode and paper mode of IELTS writing. The results pertinent to the study of Blackhurst 2005 and Weir et al. 2007 show that there were no discernible differences between

4

modes, while in one study (Green and Maycock, 2004) test-takers in the paper-based group slightly outperformed those in the computer-based one. Wolfe and Manalo (2005) came to the conclusion that there was no difference between test-takers' scores in the Test of English as a Foreign Language (TOEFL) tasks, although they had freely chosen to type or to write by hand. In a recent study, Jin and Yan (2017) stated that students in the computer-based mode attained considerably better scores than those in the paper-based mode.

There are few studies that have investigated the effect of the writing mode on the qualities of test-takers' tests. Some researchers such as Wolfe, Bolton, Feltovich, and Niday (1996) found that in computer mode the writings were mainly more formal and straighter than in the paper mode, although they were composed by the same students; but the writing mode did not influence the amount of errors for each writing significantly. Russell and Haney (1997) concluded that participants in the computer mode group were apt to compose more or less twice as much as those in the paper mode group and tended to shape their essays into further paragraphs, compared to the paper mode group.

Chambers (2008) made a comparative study of computer and paper mode groups of second language (L2) test takers. She found a greater level of lexical choice in the computer mode group, although their sentences and paragraphs were fewer. However, there was no significant difference between the two modes in terms of rate of lexical errors, length, punctuation and vocabulary use. Finally, Jin and Yan (2017) concluded that writings composed on the computer contained longer sentences and fewer errors, and they also were longer, compared to those written on paper.

The results mentioned in the previous paragraphs propose that the writing mode influences writing processes that students apply in their second language (L2) essay writing, which would in turn impact the quality of the texts they compose. In addition, differences observed in the characteristics of the essays can simply influence the test-takers' score. As an example, composing essays on the computer, as the results of some pieces of research above proposed (e.g., Jin & Yan, 2017) is connected with producing longer, richer texts in terms of lexical complexity and accuracy, and there is a probability that the essays composed on the computer attain better or higher scores than essays written on paper. This can be a description of the effect of the mode on the attained scores in L2 writing exams.

Computer familiarity and Computer Mode writing tests performance

Test-takers' familiarity with writing on the computer seems to have moderate effects on their writing performance. Torrance and Galbraith (2006) believe that from a cognitive perspective, if low-level skills such as spelling and keyboarding become automatic, then they do not need additional resources, and do not limit the writing manner. However, poor keyboarding skills redirect writers' focus of attention and cognitive re-

5

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

sources onto motor activities (i.e., typing) and, as a result, other high-order processes (e.g., revising, planning) might be left unattended to. This might result in poorer quality of the produced text (Alves et al., 2007; Horkay et al., 2006; Wolfe & Manalo, 2005). These properties could be exaggerated for L2 writers with low computer skills when they are writing in a computer exam settings (Wolfe & Manalo, 2005).

Although several studies have scrutinized test-takers' performance on paper and computer writings, only a few have surveyed the influence of computer abilities on computer mode writing performance, particularly in L2 tests (Douglas & Hegelheimer, 2007). These studies propose that test-takers with better computer familiarity seem to achieve better scores on computer mode writing, while test-takers with lower computer familiarity tend to perform better on paper mode writing. For example, Wolfe et al. (1996) claimed that writing mode does not cause a change for learners who have a good to great level of experience in computer writing in their first language (L1). Learners who had a low level of ease and practice with computers composed shorter papers with simpler sentences and obtained lower marks in computer mode than writing on paper (cf. Horkay et al., 2006; Russell & Haney, 1997).

In second language writing, Maycock and Green (2005) concluded that test-takers' skills in computer practice did not have a substantial effect on their performance on a computer-based mode of the IELTS writing. On the contrary, a study by Jin and Yan (2017) compared the writing performance of students who had similar language knowledge but different computer familiarity skills in paper and on the computer and concluded that computer skills considerably moderated the effects of writing mode on the students' writing manners and scores. Hence, students with greater computer skills were apt to achieve higher scores in computer mode. In other words, as computer familiarity improved, writing scores increased. On the contrary, participants who had lower levels of computer familiarity stated having enhanced planning while composing their writing on paper. They stressed that the computer familiarity has advantageous influences on individuals' writing style.

To sum up, there have been many studies addressing writing and the impact of using technology, specifically the computer, on the writing performance of students. Some pieces of research have been conducted on IELTS too; however, to the best of the researcher's knowledge, few studies have addressed this issue from the viewpoint of IELTS's writing sub skills. Moreover, since IELTS administrators intend to apply computer-mode testing, it is high time that research pertinent to this realm be conducted.

Purpose of the study

IELTS candidates are evaluated on four language skills: listening, speaking, reading and writing. Testing is currently in paper mode. The current research focuses on the writing section of the test, which seems to be the most demanding requirement for can-

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

didates. This study compared the performance of the participants on two different modes of IELTS Academic Writing, namely computer mode (CM) and paper mode (PM), taking into account the participants' computer familiarity. Accordingly, the following research questions were proposed:

RQ1. Do IELTS candidates perform differently in paper mode and computer mode of IELTS Academic Writing Test in terms of Task response/achievement?

RQ2. Do IELTS candidates perform differently in paper mode and computer mode of IELTS Academic Writing Test in terms of coherence and cohesion?

RQ3. Do IELTS candidates perform differently in paper mode and computer mode of IELTS Academic Writing Test in terms of lexical resource?

RQ4. Do IELTS candidates perform differently in paper mode and computer mode of IELTS Academic Writing Test in terms of grammatical range and accuracy?

Method

Participants

A total number of 108 non-native participants, who were equally categorized into PM and CM groups, were selected from 144 IELTS candidates at Tehran University (PM group = 54, CM group = 54). The participants were all adult male university students who were advanced English learners. In order to obtain proficiency-level homogeneity, all participants were selected based on their scores from the Oxford Placement Test (OPT). Although the University has categorized its language learners as advanced, the researchers administered the OPT as it had been proved reliable and valid by Cambridge Teaching English to speakers of other languages (TESOL). Thus, the participants of the study were all advanced second language (L2) learners of English who were also IELTS candidates.

Research Instruments

As was mentioned earlier, the Oxford Placement Test was administered so as to select advanced participants for the current study. Moreover, a retired IELTS Writing Academic test and a Computer Familiarity Questionnaire were administered to gather the data for further analysis.

Oxford Placement Test (OPT)
The Oxford Placement Test is a standardized proficiency test to assess students'

7

Proficiency in the English language. The OPT has been adjusted in contrast to a series of major international language examinations and can accordingly be a basis of treasured information for learners and course providers regarding appropriate course books and examination objectives, e.g. IELTS, TOEFL, TOEIC and others, such as Cambridge ESOL Main Suite exams.

Administering this test would enable the researcher to have a greater understanding of the participants' proficiency level. Hence, by administering the OPT the researchers intended to obtain participants' homogeneity regarding their proficiency level.

Normally, OPT has two sections: the first one has 40 items and students are allocated 30 to 45 minutes to do it. The second part has 20 items and the completion time allotted is 15 to 25 minutes. Only the candidates who accurately answered 36 or more questions in the first part would be permitted to move on to the next part of the test.

Retired IELTS Academic Writing Test (RIAWT)

IELTS Writing Test assesses a wide range of writing skills, including candidates' ability to provide correct answers, thought organization, and a range of vocabulary and grammar use. Examiners make use of four criteria to score each candidate: Task achievement, Coherence and cohesion, Lexical resource, and Grammatical range and accuracy. According to the test instructions, candidates should write at least 150 words in 20 minutes for Task one and 250 words in 40 minutes for Task two in academic writing. Each participant in the PM group of the study should do his writing in conventional pen-and-paper mode and should type their writing on the computer if in the CM group.

Computer Familiarity Questionnaire (CFQ)

A 14-item Computer Familiarity Questionnaire (CFQ) adopted from Weir et al. (2007) was administered to both computer and paper groups in order to find out whether or not computer familiarity had any effect on their performance. This test consisted of three categories, each of which focused mainly on a certain issue of computer familiarity, namely computer usage, comfort and perceived ability, and interest in computers. The test was based on a Likert scale from 1 (Never) to 5 (Always). The principal focus of each category is presented in Table 1.

8

Table 1. Principal focus of each category

Categories	Items
Computer Usage	Q1, Q2, Q3, Q4, Q5
Comfort and perceived ability	Q6, Q7, Q8, Q9, Q14
Interest in computers	Q10, Q11, Q12, Q13

Data Collection Procedure

The population of the study included the advanced adult participants from IELTS candidates at Tehran University. Accordingly, the OPT was administered according to the instructions of the test.

As was mentioned earlier, the OPT has two parts, comprising 40 and 20 items respectively. On the completion of the two parts of the test, the candidates' categorization was based on the following criteria (Asiyaban, Yamini, Bagheri, YarMohammadi, 2020):

- Scores 16–24————elementary
- Scores 25–40————intermediate
- Scores 43–55————advanced

According to the OPT rubric, the cut-off score for advanced learners is 41. This makes the exactness of classification between advanced and intermediate dubious. So, the researchers decided to change the cut-off point from 41 to 43. Based on the new cut-off score, only the participants who scored 43 and higher were included in the study. The test was administered in Tehran University Exam Hall under normal test conditions. The first part of the test was corrected and marked immediately in order to make sure that only the students who obtained 36 or higher would be able to continue on to the second part of the test.

The data for the study were collected at Tehran University towards the end of the term in which students had computers available to them in their classes. Students were divided into two groups, namely paper mode (PM) and computer mode (CM). Before grouping the students, the chance was given for them to choose the CM group if they were interested in being included in that group. Hence, 47 students were voluntarily put in the CM while the other 7 students were chosen randomly. In the PM group, students were given a topic (selected from retired IELTS exams) for the academic writing module and were given 60 minutes to write both tasks in the conventional paper mode. In the CM group, the students were given the same test; however, they were asked to type on the computer provided to them in their class. It is noteworthy that all proofing functions (e.g., grammar, spell check, etc.) were disabled and those who used their own laptop were double checked to ensure that the proofing functions were off. Also, all the standard conditions regarding the IELTS writing test, including acoustics, availability of necessary help and good reception, enough light, etc. were met. Finally, the participants in the CM group were also given the Computer Familiarity Questionnaire on the next session in order to determine their degree of familiarity with computers.

Reliability of the scores

Having collected the required data from the retired IELTS Academic Writing Test, the researcher then hired two trained IELTS raters to score the participants' performance in PM and CM groups. Inter-rater reliability was conducted and computed as 0.92.

9

Data Analysis

To analyze the data gathered from the instruments, SPSS package version 24 was utilized. The statistical technique adopted in the study was an Independent samples t-test. It was used to analyze the data pertinent to the performances of the participants in CM and PM group in all four band descriptors and also the data germane to the Computer Familiarity Questionnaire.

Results

The first research question examined whether IELTS candidates perform differently in PM and CM groups in terms of Task achievement. The results obtained from running the Independent samples t-test is given in Tables 2 and 3.

Table 2. *Descriptive analysis of PM and CM groups for Task response/achievement*

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task	CM	54	5.76	.88868	.12093
	PM	54	5.55	.70488	.09592

Table 3. *Independent samples t-test for two groups on Task response/achievement*

Levene's Test for Equality of Variances   t-test for Equality of Means

		F	Sig	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Task	Equal variances assumed	1.211	.274	1.380	106	.171	.21296	.15436	-.09306	.51899
	Equal variances not assumed			1.380	100.777	.171	.21296	.15436	-.09325	.51917

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

As shown in Tables 2 and 3, IELTS candidates in the CM group (M=5.76, SD=.888) did not perform differently from those in the PM group (M=5.55, SD=.704), (t (106) =1.38, p=.17, two-tailed). Thus, the answer to the first research question is negative.

To answer the second research question, stating whether candidates performed differently in paper and computer mode in terms of coherence and cohesion, another Independent sample t-test was conducted. The results are depicted in Tables 4 and 5.

Table 4. *Descriptive statistics of PM and CM groups for coherence and cohesion*

	Group	N	Mean	Std. Deviation	Std. Error Mean
Cohesion	CM	54	**6.50**	.45557	.06200
	PM	54	**4.72**	.45211	.06152

Table 5. *Independent samples t-test for two groups on coherence and cohesion*
Levene's Test for Equality of Variances   t-test for Equality of Means

		F	Sig	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Cohesion	Equal variances assumed	.168	**.683**	20.354	106	**.000**	1.77778	.08734	1.60461	1.95094
	Equal variances not assumed			20.354	105.994	.000	1.77778	.08734	1.60461	1.95094

As presented in Tables 4 and 5, IELTS candidates performed differently in paper mode (M=4.72, SD=.45) and computer mode (M=6.50, SD=.45) in terms of coherence and cohesion (t (106) =20.35, p=.00, two-tailed). The participants in the CM group significantly gained better overall band scores than those in the PM group. Hence, the answer to the second research question is positive.

The third research question explored whether IELTS candidates performed differently in paper mode and computer mode in terms of lexical resource. Accordingly, the Independent samples t-test was run. The results are shown in Tables 6 and 7.

Table 6. *The descriptive statistics of PM and CM group for lexical resource*

	Group	N	Mean	Std. Deviation	Std. Error Mean
Lexical	CM	54	**6.56**	.52347	.07124
	PM	54	**4.79**	.50017	.06807

Table 7. *Independent samples t-test for two groups on lexical resource*

Levene's Test for Equality of Variances   t-test for Equality of Means

		F	Sig	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Lexical	Equal variances assumed	.190	**.664**	17.950	106	**.000**	1.76852	.09853	1.5731	1.96386
	Equal variances not assumed			17.950	105.781	.000	1.76852	.09853	1.57318	1.96386

According to Tables 6 and 7, the mean scores for both groups are significantly different (t (106) =17.95, p=00, two-tailed). That is, IELTS candidates performed differently in the paper mode (M=4.7, SD=.500) from those in the computer mode (M=6.5, SD=.523) in terms of lexical resource. That is to say, the participants in the CM group outperformed their counterparts in the PM group. So, the answer to the third research question is positive.

The fourth research question examined whether IELTS candidates performed

differently in paper mode and computer mode of academic module of writing in terms of grammatical range and accuracy. Again, an Independent samples t-test was conducted.

Table 8. *The descriptive statistics of PM and CM group for grammatical range and accuracy*

	Group	N	Mean	Std. Deviation	Std. Error Mean
Grammar	CM	54	**6.22**	.56357	.07669
	PM	54	**5.01**	.62919	.08562

Table 9. *Independent samples t-test for two groups on coherence and cohesion*
Levene's Test for Equality of Variances   t-test for Equality of Means

		F	Sig	t	Df	Sig. (2-tai-led)	Mean Diffe-rence	Std. Error Diffe-rence	95% Confidence Interval of the Difference	
									Lower	Upper
Gram-mar	Equal variances assumed	.310	**.579**	10.472	106	**.000**	1.20370	.11495	.97581	1.43160
	Equal variances not assu-med			10.472	104.740	.000	1.20370	.11495	.97578	1.43163

As presented in Tables 8 and 9, IELTS candidates performed differently in paper mode (M=5.0, SD=.629) and computer mode (M=6.2, SD=.563) in terms of grammatical range and accuracy (t (106) = 10.47, p=00, two-tailed). Table 8 indicates that the participants in the CM group (M=6.9, SD=2.19) gained better overall band scores than those in the PM group.

Finally, to find out if CM candidates' computer familiarity differed from that of their counterparts in the PM group, another Independent samples t-test was run. The results of the t-test confirmed that the candidates' computer familiarity in the CM group was statistically different from that in the PM group. The results are presented in Tables 10 and 11.

Table 10. *The descriptive statistics of PM and CM group for Computer Familiarity Questionnaire*

	Group	N	Mean	Std. Deviation	Std. Error Mean
CFQ	CM	54	**63.94**	4.19531	.57091
	PM	54	**61.51**	5.47289	.74477

Table 10. *Independent samples t-test for two groups on CFQ*

Levene's Test for Equality of Variances   t-test for Equality of Means

		F	Sig	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
CFQ	Equal variances assumed	3.105	**.081**	2.585	106	**.011**	2.42593	.93841	.56544	4.28641
	Equal variances not assumed			2.585	99.300	.011	2.42593	.93841	.56399	4.28787

As presented in Tables 10 and 11, IELTS candidates answered the questions in the CFQ differently in the paper mode (M=61.51, SD=5.47) and the computer mode (M=63.94, SD=4.19), (t (106)=2.58, p=.011, two-tailed). That is to say, the participants in the CM group significantly gained better scores than those in the PM group which confirmed higher computer familiarity of the participants in the CM group.

Discussion

The current study investigated whether IELTS candidates performed differently in the PM and CM groups of academic writing Tasks one & two in terms of Task achievement, coherence/cohesion, lexical resource, and grammatical range and accuracy. To this end, four research questions were raised whose findings are discussed hereunder.

The findings revealed that IELTS candidates did not perform differently in PM and CM modes of academic writing tasks in terms of Task achievement. This is in support of Neuman and Baydoun's (1998) findings, who researched paper-and-pencil vs. computer testing in clerical tests. They came to the conclusion that there was no statistically significant difference between the two modes. Although no significant discrepancy was found between CM and PM, the descriptive results indicated that IELTS candidates in the CM group performed marginally better than those in the PM group in terms of Task response. They attributed such slight outperformance to participants' attention dedicated to the argumentative essay composition in Task two. In addition, as Chan et al. (2017) suggest, there is less required cognitive process in task one than in task two due to the very nature of the task. That is to say, task two is an argumentative one whose cognitive process leads to better performance on the part of the learners.

Coherence and cohesion was the second criterion based on which the performance difference of the participants of the two groups was taken into consideration. The results showed that IELTS candidates in the CM group outperformed their counterparts in the PM group, verifying the positive effects of the computer mode on writing creation. In addition, regarding Task two, the participants in the CM group gained better overall band scores than those in the PM group in terms of coherence and cohesion. This outperformance is in line with the fact that the attention that is paid to the task in the CM group is due to the nature of the task which weighs more while the screen is the platform of the work. This superiority in the CM group could also be due to the participants' familiarity with computers and also their preference for being categorized into the CM group. Although the findings of this study are opposed to those of weir et al. (2007), with regard to the influence of computer familiarity on young-adult participants' scores, this study's findings lend support to their opinion that *adults* familiar with computers tend to do better on writing tasks as they possess positive attitudes towards computers. Not only can this issue be explained in terms of writers' characteristics, but it can also be approached with regard to the physical environment (Waes, Schellens, 2003). Studying the cognitive behavior of different writers and defining five profiles for writing adoption, Waes and Schellens, (2003) concluded that the adoption of profiles depends significantly on the constraints of their writing environment.

Lexical resource was the third criterion on which the performance difference of the participants of the two groups was scrutinized. Differing from Chan et al., 2017, who concluded that students do the same in two modes of academic writing as a result of the same cognitive processes they apply, the CM group gained better scores. This result could suggest that students could review and choose better vocabularies in CM mode

15

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

which could be attributed to the results of our study, although cognitive process was not investigated.

Moreover, the results revealed that the two groups of the study performed differently in both tasks (one and two) in terms of grammatical range and accuracy. That is, the participants in CM gained better overall band scores than those in the PM group. Setting aside the tasks' nature, overall, the findings of the study are akin to Breland et al. (2004) and Wolfe and Manalo's (2005), who claimed that the computer mode is more beneficial to proficient students than to less proficient ones. As students are typing, their main focus would be on the grammaticality of the sentences and the correct choice of the forms. Therefore, they would have no concern about legibility of their hand writing. Ease of navigation was also more favored in CM group as they could easily revise and edit their writing, thus saving their time.

Moreover, the analysis of the questionnaire verified that computer familiarity seemingly influenced the candidates' scores in writing tasks in the current study. This result is also in line and opposite to the studies regarding computer familiarity and test performance. Although the findings show that the effect of computer familiarity is weighty, to the knowledge of the researchers, an advanced adult learner who is familiar and comfortable with computer usage maximizes their writing.

Furthermore, the findings of the present study are in alignment with those of Najmi's (2015) and Parsi and Sanavi's (2015) who found that utilizing technology could involve students as active learners who could adjust themselves using prompts and hints in their writings. Consequently, according to the results found in this study, assessment through technology (e.g. computer mode) in the evaluation of writing performance plays an important role in equipping participants with the tools and strategies needed to achieve optimal output. In addition, participants in the CM group manifested overall positive perception of CM which could be perceived as an influencing factor in CM participants' performance.

A number of studies have investigated the effect of the computer on L1 and L2 students' writing practices and writing quality. For instance, Shaw (2005) presented three main patterns in the findings of this line of research. First, the results are diverse, with various research finding negative impacts, some positive, and still others no effects of the computer on learners' writings. Second, the computer appears to have diverse influences on L2 writers than on L1 ones. Lastly, as most of this line of research has stressed the use of computers for educational aims, their results might have partial generalizability to assessment settings. The findings of the current study are no exception.

It seems that participants in the CM and PM groups employ different processes. According to Lee (2002), in the CM group, participants seemed to type their writing in a rough form first, then added or removed vocabulary and sentences and even paragraphs, something unmanageable in the PM group. So, it can be hypothesized that in the CM group, since revision was easier, texts could have been changed so as to avoid repetition and inappropriateness. Lee argues that some second language writers hired different

16

processes and focused on dissimilar aspects of writing across writing modes. Moreover, the candidates in the CM group seemed to be more planned which in turn confirmed the fact that they exhibited a higher level of adherence to the topic.

The computer also seems to have helped the candidates produce better texts and, as a result, receive higher scores in some band descriptors. In contrast, using a theory-based questionnaire of the writing processes, Weir, et al. (2007) showed no significant differences in terms of scores and cognitive processes across writing modes, although adult computer users who are in the high proficiency stage seem to be more used to using computers.

The matter of time in computer-assisted writing is of great importance. In spite of the belief that writing by pen, typing and looking at a screen can deviate students' focus from writing correctly, in most of the studies mentioned above, participants did not have any concern about the time allotted to doing the task. Barkaouia and Knouzib (2018) concluded that when writing on the computer, students lean towards writing considerably lengthier essays comprised of a wide selection of syntactic structures, more diverse and sophisticated lexis, more indices of local and global cohesion, than they do when writing on paper. This is not surprising since advanced learners write more frequently, perhaps on the computer, than do less advanced ones.

Overall, insights in the present study reveal those observed by Maycock and Green (2005): candidates were clearly more at ease in the computer-based mode and chose to do their tasks in the CM group since computer familiarity had made them interested and confident enough to do so. Also, candidates generally graded their computer and typing abilities as firm and outstanding in the CFQ (Q14). This is in accordance with the normally high levels of computer familiarity witnessed amongst the subjects as a whole.

6. Conclusion

In the present article the comparability of computer-based and paper-based writing tests in IELTS academic writing in terms of Task achievement, coherence/cohesion, lexical resource, and grammatical range and accuracy has been scrutinized. To achieve this aim, we looked at 108 writing samples in Iranian EFL IELTS candidates in Tehran University, Iran. The data of the study were subjected to Independent samples t-test analysis. Results showed that the students in the CM group outperformed those in the PM group in three band descriptors, namely coherence/cohesion, lexical resource, and grammatical range and accuracy.

Results also showed that computer familiarity influenced students' performance in the CM group positively. The students in the CM group were satisfied with the computer-mode test. Also, their familiarity and proficiency level worked hand in hand which led to their outperformance.

In terms of the test-taker experience, the results of the present research show that com-

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

puter-based writing was clearly preferred by CM participants, which is especially noticeable at the higher levels of English proficiency. This indicates that there is an intention towards a shift in writing on screen as being the norm, and handwriting is increasingly getting an unusual format particularly in the composition of formal and long pieces of writings; consequently, the continuance of research on the mode of test delivery is unavoidable.

Suggestions and implications

The findings of the present research might have some precious implications for policy makers. Overall, since we found a delivery mode effect in CM of IELTS writing, our results suggest that a fruitful line for further research would look more at the interaction between delivery mode, task types, and proficiency level. In a follow-up phase, it is recommended to investigate whether differences observed in writing tasks of Academic IELTS, are also manifested within the performance of the other three skills; namely Reading, Listening, and Speaking.

It should be noted, however, that the current study was performed in an academic context where candidates in the CM group had good computer knowledge, and this familiarity has improved at a fast rate in recent years. Computer familiarity and practices concerning paper versus computer methods of writing may be immensely diverse in some other settings, with potential consequences for test delivery mode effects. Hence, supplementary investigation is required to approve the generalizability of the findings of this study and the scope for online IELTS modes of delivery worldwide.

18

References

• Asiyaban, A. R., Yamini, M., Bagheri, M. S., &Yarmohammadi, L. (2020). Implicit/explicit knowledge and its contribution towards tense consistency employment across EFL learners' proficiency levels. *Cogent Education*, 7(1), 1727129.

• Alves, R. A., Castro, S. L., de Sousa, L., &Stromqvist, S. (2007). Influence of typing skill on pause-execution cycles in written composition. In M. Torrance, L. van Waes,& D. Galbraith (Eds.). Writing and cognition: Research and applications (pp. 55–65). Amsterdam: Elsevier.

• Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. Mod. Lang. Rev. 100(1), 320–340.

• Barkaoui, K., &Knouzi, I. (2018). The effects of writing mode and computer ability on L2 test takers' essay characteristics and scores.*Assessing Writing*, 36, 19-31.

• Blackhurst, A. (2005). Listening: Reading and writing on computer-based and paper-based versions of IELTS. Research Notes, 21, 14–17.

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

• Chambers, L. (2008). Computer-based and paper-based writing assessment: A comparative text analysis. Research Notes, 34, 9–15.

• Chan, S., Bax, S., & Weir, C. (2017). Researching participants taking IELTS Academic Writing Task 2 (AWT2) in paper mode and in computer mode in terms of score equivalence, cognitive validity and other factors.

• Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Ernst KlettSprachen.

• Clariana, R., & Wallace, P. (2002). Paper–based versus computer–based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology,* 33(5), 593-602.

• Davies, A. (2007). Assessing academic English language proficiency: 40+ years of U.K. language tests. In J. Fox, M. Wesche, D.Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 73–86). Ottawa: University of Ottawa Press.

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

19

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe
Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

• Green, T., &Maycock, L. (2004).Computer-based IELTS and paper-based versions of IELTS. Research Notes, 18, 3–6.

• Fulcher, G. (2014). Philosophy and language testing. In A. J. Kunnan (Ed.), *The companion to language testing* (pp. 1431–1451). Hoboken: Wiley.

• Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP.Journal of Technology, Learning and Assessment, 5(2), Retrieved from: www.jtla.org.

• IELTS. (2014). Guide for educational institutions, governments, professional bodies and commercialorganisations. Retrieved from https://www.ielts.org/- /media/publications/ guide-for-institutions/ielts-guide-for- institutions-uk.ashx?la=en

• Jin, Y., & Yan, M. (2017). Computer literacy and the construct validity of a high-stakes computer-based writing assessment. Language Assessment Quarterly, 1–19.

• Lee, Y. J. (2002). A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessing Writing*, 8(2), 135-157.

• Li, J. (2006). The mediation of technology in ESL writing and its implications for writing assessment. *Assessing Writing*, 11(1), 5-21.

• Lottridge, S., Nicewander, A., Schulz, M. &Mitzel, H. (2008). Comparability of Paper-based and Computer-based Tests: A Review of the Methodology.Pacific Metrics Corporation 585 Cannery Row, Suite 201 Monterey, California 93940.

• Maycock, L. O. U. I. S. E., & Green, T. O. N. Y. (2005). The effects on performance of computer familiarity and attitudes towards CB IELTS. *Research Notes*, 20, 3-8.

• Najmi, K. (2015). The effect of genre-based approach on enhancing writing skill of Iranian law students. *Modern Journal of Language Teaching Methods*, 5(2), 474.

• Neuman, G., &Baydoun, R. (1998). Computerization of paper-and-pencil tests: When are they equivalent?. *Applied Psychological Measurement,* 22(1), 71-83.

• O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they?. *Language Assessment Quarterly*, 8(2), 146-160.

• Parsi, P., &VahdaniSanavi, R. (2015). The effects of dynamic assessment on improving writing ability of intermediate EFL learners. *International Journal of Language Learning and Applied Linguistics World*, 8 (2), 73-88.

20

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

• 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), 5-30.

• Puspawati, I. (2012). EFL/ESL (English as a Foreign/Second Language) Students' Perceptions toward the TOEFL (Test of English as a Foreign Language) Test.

• Quaid, E. D. (2018). Reviewing the IELTS speaking test in East Asia: Theoretical and practice-based insights. *Language Testing in Asia,* 8(2), 25-48.

• Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. Education Policy Analysis Archives, 5(3), Retrieved from: http://epaa.asu.edu/epaa/v5n3.html.

• Shohamy, E. G. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education.

• Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham,& J. Fitzgerald (Eds.). Handbook of writing research (pp. 67–80). New York: Guilford Press.

• Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT journal,* 64(3), 314-320.

• Van Waes, L., &Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of pragmatics*, 35(6), 829-853.

• Weir C J, O'Sullivan B (2017) 'Assessing English on the global stage: the British Council and English language testing, 1941-2016', in (ed(s).). , edn, London: Equinox.

• Weir, C. Y. R. I. L., & Shaw, S. T. U. A. R. T. (2005). Establishing the validity of Cambridge ESOL writing tests: towards the implementation of a socio-cognitive model for test validation. *University of Cambridge ESOL Examinations Research Notes,* 21, 10-14.

• Weir, C., Yan, J., O'Sullivan, B., &Bax, S. (2007). Does the computer make a difference?: The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact. *International English Language Testing System (IELTS) Research Reports 2007: Volume 7*, 1.

• Weir, C. J., O'Sullivan, B., & Jin, Y. (2007). Does the computer make a difference? The reaction of test-takers to a computer-based versus a traditional hand-written form of the IELTS writing component: Effects and impact. IELTS Research Report No.7British Council & IDPAustralia311–347.

21

Literatura / Filosofía / Arte
Historia / Libros / Seminarios

álabe

Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red de Universidades Lectoras Revista de la Red d

•Wolfe, E. W., & Manalo, J. R. (2005). An Investigation of the Impact of Composition Medium on the Quality of TOEFL Writing Scores. TOEFL® Research Report.RR-72.ETS RR 04-29. *ETS Research Report Series.*

•Wolfe, E. W., Bolton, S., Feltovich, B., &Niday, D. M. (1996). The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessing Writing*, 3, 123–147.

•Yurdabakan, I. (2012). Primary School Students' Attitudes Towards Computer Based testing and Assessment in Turkey. *Turkish Online Journal of Distance Education*- 13 (12), 177-188.

22