Grammatical accuracy & length across computer-assisted and pencil-and-paper assessments: Variation in advanced-level EFL writing

Massoud Yaghoubi-Notash (PhD), Maryam Mahmoodi
English department, Faculty of Persian Literature & Foreign Languages,
University of Tabriz, Tabriz, IRAN. P O BOX 51835331
masoud.yaghoubi@gmail.com
Aras International Campus,
University of Tabriz, Tabriz, IRAN.
m.mahmoodi_1392@yahoo.com

Abstract
The sweeping trend of communication technologies increasingly makes computers and computer-based operations an indispensable part of our everyday lives. Education as an institution deeply rooted within the context of the broader community can, therefore, hardly afford to do without the digital era. Assessment, as a subcategory of education, is expected to be under the influence of the same trend. The present study, focusing on the EFL situation in Iran compared two equal-sized advanced learners’ each containing 20) writing accuracy and length. One group employed the traditional paper-and-pencil (PPT) mode while the other used computer-assisted language testing (CALT). Results indicated a significantly more accurate writing for the CALT group, while a longer writing production for the PPT group. Implications of the study are discussed.

Key Words: CALT, grammatical accuracy, length; PPT

Academic Discipline And Sub-Disciplines:
EFL education
SUBJECT CLASSIFICATION
Assessment, Computer-assisted language testing (CALT)
Introduction

Writing as a means of communication occupies a central position in the professional and personal spheres of human life. In addition to being a vehicle of communication for formal and official correspondence, it is so closely associated with the academic skills. In the latter vein, it requires not only the mastery of language skills (e.g., the right style, genre, etc.), but also a knowledge of content. In being so, it is supposed to be quite a time-consuming activity (Dastjerdi & Samian, 2011). Kashani, Binti Mahmoud & Rezvani Kalajakhi (2013, p. 204) state that

Through making use of language's depth and originality [in writing], the personal growth process is recorded by individuals. This enriches their lives and vitalizes the employment of language. Composition can also be regarded as a combined representation of all four language skills (consisted of writing, reading, speaking, and listening) and can be an appropriate means of reflecting the overall language competence and proficiency. But, due to the complex nature of the composition's content, it is a particularly difficult subject for students.

It can also be argued that all four skills are involved interactively in composition writing, and for this very reason can be a good source of judgment as to the learners' overall language competence and proficiency. According to Obemeata (1995), English as a foreign language functions as a system of communication which is superimposed upon an already existing medium of communication, that is the first language. Also he argues that writing and reading are demanding and take more effort for instruction than listening and speaking in that the latter develops naturally like a growing child.

It can be argued that writing is the most complicated skill of all that necessitates complicated cognitive activity and a simultaneous control of all variables (see Kolade, 2012). Here are some of the factors that are presumably effective and helpful in writing improvement:

- Awareness of the students' needs (Tribble, 1996);
- Purposefulness of the writing and alignment with the students' interest (Bryan, 2004);
- Quality enhancement of the thinking process on the students' part (Barra, 2005);
- The possibility of drafting and self-revision (Doughty and Pica, 1986);
- Emphasizing the process of writing which encourages autonomy (Vanessa, 2004).

Computer-assisted writing in ESL

Computers were used in the classrooms as early as the 1980s when the writing teachers in American schools and universities applied word-processing to facilitate students' writing outputs. Kulik (2003) indicates improvements in the performance of students using a word processor. Despite the fact that findings do not lend a high support for impacts in the majority of word-processing researches, the impacts were still large enough to be viewed as educationally meaningful (ibid). As Pennington (2003) declares, writing in a computer context is more advantageous for students than writing by pen-and-paper due since the former is more practical and convenient and quite flexible. However, according to Kern and Warshauer (2000) computer, in the same vein as any other technological device does not generate advancements in learning in and of itself. On the other hand, Lee et al. (2009) believe that EFL writing teaching and learning problems can be reduced by using computer technologies. Traditionally, computers were used as a mere device or instrument; nevertheless, nowadays Computer Assisted Language Learning (CALL) activities such as e-mails and blogs have replaced the old function of language in learning (Pennington, 2004).

According to Kupelian (2001), an efficient educational resource, the electronic tool has not only changed the composing process, but also strengthens participation in writing activity. He goes on to lay emphasis on the non-threatening atmosphere provided by the computers in which writers get encouraged to be productive in terms of language. Another reason is that the Web provides a field for writers to present their work to a real and larger audience that spreads beyond classroom and school frames (Karchmer, 2001). One of the e-mediums which can be used in language learning and teaching writing is weblog or blog. According to Bella (2005), a weblog, also called a blog, is an easily created and updateable website that allows people to publish to the Internet instantly. Soares (2008) asserts that the blogs because of their public and interactive aspects can be applied for educational purposes such as students' interests and class content. When students realize that they put their work on the weblog for readers in the real world, they are motivated to write (Leibowitz, 1999). Ward (2004) sees blogs as useful and helpful for language learning and teaching.

As Shi-Jer Lou et al. (2010) indicate, blogs enable students to interact and cooperate with each other. Yang and Chen (2007) believe a number of benefits of blogging applications in instruction to be: knowledge creation, the construction of learning files, enhancement of writing skills and reduction of misspelling, enhancement of school reputation and administration efficiency, and the enhancement of the exchange between the instructors and students. It has also been indicated by Du and Wagner (2007) that these blogs allow individuals to make their thoughts public in web pages and thus share their knowledge and thoughts. Because of the potential effect blogging has on the sharing and expressing of knowledge, blogging has a positive impact on learning and instruction (Brownstein & Klein, 2008; Dippold, 2009; Goldman, Cohen & Sheahan, 2008). Warlick (2005) suggests that

When students are writing large reports, essays, or research papers, ask them to submit their rough drafts onto their blogs. Then comment on the papers constructive criticisms. This is not to say that personal, face-to-face assistance is not needed. The advantage of using a blog is that all comments and developments are archived for study and reference” (p. 154).

Halic, Lee, Paulus and Spense (2010), in this concern, have adeptly noted that:
The popularity of blogs among young people has made the map pealing to educators seeking to integrate computer-mediated communication (CMC) tools at the university level. These tools are seen as having the potential for enhancing student engagement and providing an environment for collaboration and creation of knowledge (p. 1).

In the same vein, Campbell (2003) discusses the potentiality and practicability of integrating weblogs into the educational context, particularly in the field of language teaching.

A study was conducted by Zare-ee, Shekarey and FathiVajargah (2009) among the Iranian undergraduate and graduate students in Iran on the use of technology in education. In their study, students’ views on the application of blogs in teaching-learning processes and their actual use of blogs were investigated. The participants’ familiarity with blogs which were written in English and Persian was studied and their views about the usefulness of blogs as an instructional tool in higher education were declared. The findings of the study showed that blogs were actually used less than it is believed they should be. It was one of the studies done in Iran related to the use of the blog in education at university level. The study did not investigate the students’ writing performance and was aimed to view the students’ perceptions about blog in teaching-learning process. Also, the research methodology of that study was qualitative survey and not experimental.

Pen-and-Paper versus Technology-Based Writing

Lam and Pennington (1995) conducted a study in which they compared two groups of secondary students who used pen-and paper and computer for writing. Their writings were assessed by raters based on ESL Composition Profile (1981) and the results presented that that overall, the writing of students in the Computer group was better than that of the one in a Pen group and they were hugely significant differences in every aspect of their writing except organization and content with the superior exhibition of performance belonging to the Computer group.

A few studies indicate the positive effect of using weblogs to enhance students’ writing development. Ward (2004) reported that students enjoyed the experience of writing in a blog with signs of improvement due to a high level of motivation. Though they are not intended for language learning and education purposes, blogs seem to have a great potential for EFL pedagogy (Kavaliauskienė & Mažekioniene, 2006). However, most of the researchers in this field believe that the full potential of writing class blogs still needs to be learned and explored and also a lot of work needs to be done in order to effectively use blogs in the writing class. Aside from other pedagogical merits, writing improvement and enhancing motivation, there are other effects on learners such as fostering autonomy, peer and group work and feedback provision.

Similar findings were reported in studies by Russell (1999) and Russell and Plati (2001), who showed that students who were accustomed to writing using a computer performed better on the open-ended tests when they wrote using a computer than when they wrote by hand. Pommerich (2004) investigated the item-level mode effects of English, reading, and science reasoning tests in grades 11 and 12 and found that the mode effect in general was small. Two comparability studies on the online versions of the NAEP math and writing tests showed that the paper group significantly outperformed the computer group in the eighth-grade NAEP mathematics test, but no mode effect was found for the eighth-grade NAEP essay test (Sandene, Horkay, Bennett, Allen, Braswell, Kaplan, 2005). The NAEP studies also found that students’ familiarity with computers was related to their performance. Particularly, hands-on measures of keyboarding skill were found to be a significant predictor of students’ performances on the NAEP online writing test (Sandene, et al., 2005). Experimenting the effect of different modes i.e. conventional and computer-based) on children's performance, Barnes (2010) could prove a significant improvement for children; however, a lot seemed to depend on the content and age. With this background in mind, the present study seeks to answer the following research questions:

R.Q. 1. Does advanced EFL learners’ writing accuracy vary significantly across pencil-and-paper and computer-assisted testing?

R.Q. 2. Does advanced EFL learners’ writing length vary significantly across pencil-and-paper and computer-assisted testing?

METHOD

Participants

Participants in the study were two equal-sized groups each consisting of 20 all male students. Both groups were in fact classes in an institute in Shabestar, a township in Eastern Azerbaijan Province in Northwest of Iran. Their ages ranged between 22 and 32. They had already passed conversational courses for at least 10 semesters and were currently doing their advanced level spoken English courses. As participants, they were given Oxford Placement Test based on which they were ranked as advanced level students. Their first language was Azerbaijani as they belonged to the local context of the township.
Instrumentation

The instruments of the study were as follows:

- Oxford Placement Test Packages: The Oxford Placement Tests provide a reliable and efficient means of placing students at the correct places.

The new edition consists of:

Test packs 1 and 2 containing redesigned test papers with revised and updated items,

- A Marking Kit with User's Guide and Diagnostic Key, containing marking overlays, detailed guidance on administering and grading the Tests and diagnostic grammar references,
- Cassettes for Tests 1 and 2 with revised and updated test items.

Each Oxford Placement Test Pack consists of two test pads containing enough copies of the Listening and Grammar Tests for students, the Teacher's introduction and a key for the tests. A C10 cassette accompanies the Listening Test. The Oxford Placement Tests can be used with any number of students of English to ensure efficient, reliable and accurate grading and placing of students into classes at all levels from elementary to post-proficiency. They can determine realistic examination targets for groups or individuals.

The test is divided into two main sections (Listening Test and Grammar Test), each of 100 items. The first section is primarily a test of reading and listening skills, in which the learner's performance is dependent on knowledge of the sound and writing system of English and on the ability to make use of this knowledge at a task-speed well within the competence of a native speaker of English. The second section is a test of grammatical structure and involves a carefully selected range of those structures consistently found in course books and examinations at elementary, intermediate and advanced levels.

The two sections of the test are, however, designed to be used together to produce a total score out of 200. It is this aggregate score from the two sections that should ideally be used to establish a rank order for placement or other purposes.

A written key to the test appears on the inside cover of the Test Pack. The Marking Kit with User's Guide and Diagnostic Key contains a separate marking overlay for each page of the test. The Levels Chart shows levels relationships which represent broad statistical correlations. It, therefore, offers a very useful general guide to where learners are on a number of widely recognized scales of overall language proficiency

- White Smoke software

White Smoke is an English writing tool that provides grammar, spelling, punctuation and style checks. Integrated into White Smoke are world-renowned word and text translation and document templates. White Smoke is activated in a single click from any text application and browser. It is an added-value product that ensures a higher standard of English writing. White Smoke encourages the writing process, calls out potential errors for consideration, and offers suggestions for improvement.

Procedures

There were 315 EFL learners studying English in padideh Language Center, Shabestar. The language proficiency level of 100 learners had been determined through an Oxford Placement Test by the authorities of the institute, in advance.

In the main study, 20 participants out of 100 intermediate EFL learners were selected non-randomly.

At the beginning of the study, the participants were engaged in traditional writing, which is the participant wrote essays about the topic (what can be an ideal marriage in your country) on the paper and with pen or pencil during an hour. While, after two weeks the participants were engaged in modern writing. They typed their essays on the computer by keyboard about the same topic and during an hour, too.

In the PPT performance, papers were scored by a rater according to the grammar errors on the papers. Also the length of the essays was considered (the number of complete sentences in an essay was counted).

In the computerized system, essays were scored automatically by software that was connected on to the Internet according to the grammar errors.

DATA ANALYSIS & RESULTS

In order to analyze the data, the researcher used both descriptive as well as inferential statistics. A series of SPSS were performed on the data whether there were any significant differences in performance across the two modes. To answer the research questions the researcher conducted a series of paired samples t-tests to observe how each group performed independently and comparably. And also the researcher used rank order analysis for the comparison of the performance of two modes, CALT and PPT. The details of the statistical analyses, the results, and the outcomes of the study are presented and discussed in the following lines.

The descriptive data of the results for placement test OPT), grammatical accuracy, and length are presented in Tables 1 to 3.
As can be seen, Table 1 shows that the mean of students on OPT (Oxford Placement Test) is high enough for the researchers to place them at the advance level of proficiency.

Table 1. OPT results descriptive data

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Sum</th>
<th>Mean</th>
<th>Variance</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPT</td>
<td>20</td>
<td>3477</td>
<td>173.85</td>
<td>109</td>
<td>10.44</td>
</tr>
</tbody>
</table>

After administering the writing test in PPT, the mean turned out to be 69, whereas it was 82.50 for CALT.

Table 2. Descriptive data of grammatical accuracy assessment using PPT and CALT methods

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Sum</th>
<th>Mean</th>
<th>Variance</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT*</td>
<td>20</td>
<td>1380</td>
<td>69</td>
<td>183.15</td>
<td>13.53</td>
</tr>
<tr>
<td>CALT**</td>
<td>20</td>
<td>1650</td>
<td>82.5</td>
<td>85.52</td>
<td>9.24</td>
</tr>
</tbody>
</table>

* PPT = Pencil-and-paper test; ** CALT = Computer-assisted language test

As can be seen in Table 3 the mean length in PPT exceeds that of CALT which is quite diametrically opposed to the difference regarding grammatical accuracy across the two modes.

As for the first research question, "Does advanced EFL learners' writing accuracy vary significantly across pencil-and-paper and computer-assisted testing?", the results were analyzed employing paired-samples t-test. The inferential statistics appears below in Table 4.

Table 3. Descriptive data of writing length using PPT and CALT methods

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Sum</th>
<th>Mean</th>
<th>Variance</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPT*</td>
<td>20</td>
<td>1505.68</td>
<td>75.28</td>
<td>75.24</td>
<td>8.67</td>
</tr>
<tr>
<td>CALT**</td>
<td>20</td>
<td>1085.66</td>
<td>54.28</td>
<td>127</td>
<td>11.27</td>
</tr>
</tbody>
</table>

* PPT = Pencil-and-paper test; ** CALT = Computer-assisted language test

Regarding the grammatical accuracy, the results clearly illustrated that the mean difference (13.5) was simply significant at p < 0.05. Therefore the mean of the scores across the two sets of tests were numerically distanced enough beyond random variation.

Figure 1. Means of grammatical accuracy in PPT and CALT methods
Having made sure that the difference is significant at \( p < 0.05 \), a look back at the mean can clearly illustrate that the mean of the CALT group is higher than that of the PPT group. Therefore, the CALT group outperformed the PPT group regarding grammatical accuracy.

For the second research question, namely "Does advanced EFL learners’ writing accuracy vary significantly across pencil-and-paper and computer-assisted testing?" the results are presented in Table 5.

<table>
<thead>
<tr>
<th>Mean Difference</th>
<th>Observed t-value</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>7.10</td>
<td>19</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The calculated \( t \)-value, 7.10, was found to be significant with \( p < 0.05 \) thus, rejecting the null hypothesis (Table 4.10). The research indicated that there was a significant difference between subjects’ length performance on paper-and pencil test and computer-assisted test.

**Figure 2. Means of writing length in PPT and CALT methods**

Now that the mean difference between the two groups i.e. PPT & CALT) was statistically significant, comparing the means in terms of their numerical value clearly indicates that the PPT group outperformed CALT group in length.

**DISCUSSION & CONCLUSION**

As far as the first research question is concerned, the difference in terms of grammatical accuracy across Paper-and-Pencil Test (PPT) and Computer-assisted Language Testing (CALT) turned out to be significant. The participants who had taken the writing test through CALT had produced significantly more accurate writings than their PPT counterparts. This finding is in line with Barnes (2010), Kulik (2003), Lam and Pennington (1995), Pommerich (2004), Russell (1999), Russell and Plati (2001), Sandene, Horkay, Bennett, Allen, Braswell, and Kaplan (2005). This finding is justified by Dautie (1986) who suggested that computer-based writing offers the possibility of revision which may be a source of grammatical accuracy. However, the finding fails to be supported by Kern and Warshauer (2000).

Concerning the second research question PPT group outperformed CALT which means that using the traditional paper and pencil mode prompted the students to write more in terms of length. This finding is in line with the earlier findings such Kern and Warshauer (2000), but is contradicted by Barnes (2010), Kulik (2003), Lam and Pennington (1995), Pommerich (2004), Russell (1999), Russell and Plati (2001), Sandene, Horkay, Bennett, Allen, Braswell, and Kaplan (2005).

This study bears implications for language assessment where CALT can offer great potentials for gauging learner performance. Educational systems at large that are facing a sweeping trend of electronic and digital developments can benefit from CALT and its vast array of data. Also, material and syllabus designers, language teaching practitioners, and many other local pedagogical concerns with language learning and teaching.
References


