

The Journal of Teaching Language Skills (JTLS)
3(2), Summer 2011, Ser. 63/4
(Previously *Journal of Social Sciences & Humanities*)

An Investigation into Bilingual Dictionary Use: Do the Frequency of Use and Type of Dictionary Make a Difference in L2 Writing Performance?

A. Roohani *

Assistant Professor

Shahrekord University, Shahrekord

email: roohani.ali@gmail.com

A. Khosravi

M. A., TEFL

Islamic Azad University, Najafabad

email: amineh.khosravi@gmail.com

Abstract

Bilingual dictionary use in L2 writing test performance has recently been the subject of debate. Opinions differ according to how the trait is understood and whether the system favors the process-oriented or product-oriented views towards the assessment and writing skill. Given the need for more empirical support, this study is aimed at investigating the availability of bilingual dictionary use, frequency of use and type of bilingual dictionary (Persian-to-English and/or English-to-Persian) in L2 writing tests. To these ends, 60 Iranian EFL learners, in a counterbalancing procedure, were asked to write 2 timed essays in 2 test conditions: with and without a dictionary. The frequency of use in this study was defined in terms of number of look-ups in the bilingual dictionaries. The t-test between writing mean scores in the 2 test conditions revealed that using a bilingual dictionary made a significant difference in the writing test scores. The results of ANOVA also showed that frequency of dictionary use made a significant difference, but the high frequency of use was not found to be positive. Furthermore, the type of bilingual dictionary was found to be a significant variable, with the highest writing mean score for those who used English-to-

Received: 07/06/2010 Accepted: 03/01/2011

* Corresponding author

Persian dictionaries in writing the essays. Finally, pedagogical implications for L2 learners and teachers are presented.

Keywords: bilingual dictionary, frequency of use, type of dictionary, writing

1. Introduction

Writing, as Chastain (1988) states, is a basic communicative skill and a unique asset in the process of learning. Widdowson (1983) considers it an interactive process of negotiation. However, providing a cohesive and coherent piece of writing is difficult since, as Zamel (1987) states, writing has a complex, recursive and non-linear nature requiring a variety of microskills. This might be a reason why there is no consensus among second language researchers over the best approach to teach or learn it. However, as Raimes (1990) points out, there are two major lines of thinking with regard to the writing skill: Product-oriented and process-oriented views. Advocates of writing as product emphasize the finished product in terms of grammar, vocabulary, spelling, and cohesive devices whereas advocates of writing as process focus on what happens when a writer is writing; that is, the process of writing which leads the writer to generate ideas and organize them into a coherent sequence

The above issue has been fuelled by the role of dictionaries in the language assessment. Whether dictionaries —monolingual or bilingual— should be used as a tool in L2 writing tasks/tests has been the subject of debate for several years (see East, 2008; Hartman, 1990; Horsfall, 1997; Thompson, 1987). The debate around assessment practice exemplifies the conflict between two opposing approaches towards writing, depending on how the construct is understood: Whether the value system favors product-oriented tests and emphasizes the discriminatory power of the tests, or whether it supports process-oriented tests and is just concerned with learners' language performance. As East (2008) states, the use of dictionaries might be less favored if test developers are concerned with product-oriented tests because, in Bachman's (1990) terms, the primary function of tests is that of a request for information about the test takers' language ability, without allowing anything which

might change test takers' performance. On the other hand, the use of dictionaries might be more favored if language teachers are concerned with the process-oriented assessment. As Chambers (1999) points out, using bilingual dictionaries in the process of L2 writing can be authentic and contribute to test takers' perceived relevance. According to East (2008), this perceived relevance helps to promote a positive affective response to test tasks, hence helping test takers to perform at their best.

EFL teachers in Iran, perhaps like other L2 teachers in other countries, might be in dilemma to allow or ban their students to use bilingual dictionaries in writing tasks/tests. Some ban its use as it might make the test too easy in a way that it might not be relevant to the writing construct and, as such, test scores will not be quite valid. Others allow it as bilingual dictionaries may potentially improve the quality of the product. Furthermore, EFL teachers are undecided on the type of bilingual dictionary and frequency of dictionary use. If test takers are allowed to use bilingual dictionaries, they can use L1-to-L2, L2-to-L1 or both types of dictionaries; which one represents the construct more validly and offers more benefits to L2 learners is the subject of investigation. Unfortunately, few studies have put the effect of type of dictionary on L2 writing performance under their spotlight. Meanwhile, as Elola (2008) states, the applications of dictionaries and strategies used by L2 learners in L2 writing tasks are varied. It is likely that very high or low degree of frequency of dictionary use obscures the information about the test takers' language ability from the test. Given the above issues, this study is designed to clarify the role of using bilingual dictionaries, frequency of dictionary use and type of bilingual dictionary in L2 writing performance of Iranian EFL learners.

2. Review of Literature

Some part of literature on EFL/ESL writing (e.g. Atwell, 1985; Henry, 1996; Hillocks, 1986) emphasizes writing as product, that is, the visible outcome providing us with the manifestation of the knowledge possessed by the writer as the result of activating the knowledge of vocabulary,

grammar, and spelling. Hence, L2 teachers drill the various grammar forms and correct use of forms, expecting L2 learners to develop the skills required to communicate the message. Henry (1996) points out that most studies on L2 writing conducted by native speakers of English also focus on product such as the effects of word-level or sentence-level error correction.

On the other hand, a large body of literature on L2 writing (e.g. East, 2008; Kern & Schultz, 1992; Zamel, 1987) challenges the product approach and, instead, focuses on the process-oriented view. According to Zamel (1983), researchers have realized that the investigation of students' written products do not reveal much about their instructional needs. That is why they are now exploring writing behaviors, which can provide insight into how to teach it. In this view, what is required is the establishment of a supportive environment in which L2 learners are encouraged to take risk and get engaged in creating meaning.

Likewise, the debate surrounding assessment practice is indicative of the conflict between two main approaches towards writing. As East (2008) states, one view is influenced by knowledge-based approaches that favor the *static* assessment of writing. The advocates of this view use tests that produce a snapshot of the L2 test takers' writing ability and measure knowledge of key components. They focus on the discriminatory power of the test to predict future success. The other view is influenced by the process-oriented approach that focuses on *dynamic* assessment, which is more learner-centered. According to this view, communicative writing proficiency construct is as an authentic reflection of writing as process. Thus, a dictionary can be an authentic tool in the process of writing process.

In the same manner, dictionary use —monolingual and bilingual— in L2 tests has been a contentious issue, sometimes leading to contradictory claims. Thompson (1987) states that, unlike bilingual dictionary use, monolingual dictionary use by L2 learners seems to be the accepted orthodoxy. However, the study by Martin (2002) showed that bilingual dictionary use is beneficial. He carried out a study with twenty

native Chinese speakers to show the helpfulness of monolingual dictionaries, compared with bilingual dictionaries, in written English when dealing with transitive and intransitive verbs. The participants were divided into 2 groups and given a survey regarding their knowledge and attitudes toward dictionaries. One group was given a bilingual dictionary and the other group received a monolingual dictionary while both of them were presented a brief lesson on transitive and intransitive verbs. The results proved the importance of a bilingual dictionary with 85% of the participants making use of such dictionaries.

The results of other research into dictionary use in L2 tests seem to be inconsistent. For instance, Benssoussan, Sim and Weiss (1984) investigated the effects of dictionary use on L2 learners' reading tests performance. The results did not show any significant difference between mean scores of those who used the dictionary and those who did not. Yet, Hayati (2005) has reported that dictionary use makes a difference in L2 reading performance. He investigated the impact of bilingual dictionaries, compared with monolinguals, on reading comprehension of forty-five intermediate L2 learners. The participants were divided into three equal groups: bilingual group (BG), monolingual group (MG), and control group (CG), and took the same reading comprehension test. BG used only bilingual dictionaries, MG used only monolingual dictionaries and CG took advantage of guessing ability and contextual cues with no reference to any kind of dictionary. The results indicated that BG received the highest mean scores, indicating that bilingual dictionaries could be a powerful pedagogical tool.

As to the writing, Engber (1995) carried out a study on the impact of bilingual dictionaries on writing quality. She examined the extent to which measures of lexical richness and lexical accuracy were related to the quality of the writing products of intermediate level learners of English. Her findings indicated that bilingual dictionary use could affect lexical accuracy and lexical range, which were two components of the writing product. Also, Atkins and Varantola (1998) had a large-scale investigation with L2 learners drawn from a range of educational

institutions across Europe. They investigated dictionary use among learners with different language abilities while they were doing a set of tasks including writing tasks. They found out that the lower ability participants of the study made use of dictionary more than the higher ability participants of the study, and the most proficient participants used the dictionary the least.

As to the writing tests, Hurman and Tall (1998; cited in Tall & Hurman, 2002) carried out a major study with groups of French students in the UK. This substantial study used a comparative within-subjects design whereby the participants ($n = 1300$) were asked to take two writing tests in French, one 'with' and one 'without' a dictionary. The researchers discovered that the participants gained an average increase of two marks (9%) on the 'with dictionary' paper in comparison to their scores on the 'without dictionary' paper. The results showed that the test scores were significantly higher when dictionaries were available. This study had a considerable consequence for advocating dictionary use in testing practice in the UK. However, it had some limitations in the design. For example, the 'without dictionary' tests were mainly administered before the 'with dictionary' tests, which led to practice effect. Besides, both tests were scored with the same rater inconsistently, which led to intra-rater error.

Also, East (2005) conducted a study in the context of student participation in an intermediate level German course in a tertiary institution in New Zealand. The study was part of a series of investigations into bilingual dictionary use in writing tests. He investigated the extent to which dictionary use in the testing context impacted in two areas: whether, on the one hand, the dictionary interfered with the test as a measure of test taker ability as reflected in the scores; and whether, on the other hand, the use of the dictionary as a tool for independent learning could be transferred to the testing context in a way that enhanced both the quality of what test takers wrote and their positive experience of the tests. The results indicated that use of the dictionary made no significant difference to final scores, irrespective of the type of

writing test (i.e., with or without a dictionary). Also, the findings suggested that, on the one hand, the availability of the dictionary as a support resource was psychologically motivating, but, on the other, tailor-made resources could be of more value.

In another study, East (2006) investigated whether allowing a bilingual dictionary would enhance good writing in German writing tests. The participants ($n = 47$) took two tests: one with and one without a bilingual dictionary. The essays were compared in terms of test scores, lexical sophistication and lexical accuracy, defined through a quantitative analysis of words looked up in the dictionary. The results revealed an increase in lexical sophistication in 'with dictionary' tests with frequent misuse of look-ups and no improvement in test scores. Finally, East (2007) addressed the frequency of dictionary use in the writing performance of students studying German in New Zealand. The results of the multiple regression analysis demonstrated that the frequency of dictionary use, operationlized in terms of number of look-ups, failed to significantly predict performance on the 'with dictionary' writing test performance.

The discrepant findings on availability of dictionary use alone indicate that further research is required to better establish the extent to which availability of dictionary to EFL learners has a differential effect. Given the debate around dictionary use in testing and inconsistency of results, any decision made on the basis of little research might be unwise. Besides, the lack of empirical research into the type and frequency of bilingual dictionary use warrants further research to shed more light on the construct of L2 writings. This study therefore seeks to address these issues.

3. Objectives

This study is intended to investigate whether the use of bilingual dictionaries has a differential effect on the L2 writing performance of EFL learners in Iran. One way to provide essential information to probe the purpose of this study, as Rivera and Standsfied (1998) suggest, is to

make use of test scores information through comparative within-subjects design. That is the reason why this study makes use of two conditions: with and without a dictionary. Meanwhile, the use of monolingual dictionaries has been excluded from this study since, as East (2008) states, the controversy over the monolingual dictionary use in testing has recently been shifted to the bilingual dictionary use. Besides, adding one more variable could make the design and interpretation of results more complex.

Additional purpose of this study is to explore the effect of type of bilingual dictionaries (L1-to-L2, L2-to-L1 or both) and frequency of dictionary use on the L2 writing performance. Accordingly, the following research questions are investigated:

1. Does the bilingual dictionary use make a significant difference in test takers' L2 writing performance?
2. Does the frequency of dictionary use make a significant difference in test takers' 'with dictionary' writing test performance?
3. Does the type of dictionary (English-to-Persian, Persian-to-English or both) make a significant difference in test takers' 'with dictionary' writing test performance?

The following three null hypotheses are addressed in this study:

H₀₁: Using bilingual dictionaries makes no significant difference in test takers' L2 writing scores.

H₀₂: The frequency of dictionary use (high or low) does not make a significant difference in test takers' 'with dictionary' writing test performance.

H₀₃: There is no significant difference among 'with dictionary' writing test scores of the EFL participants who use English-to-Persian, Persian-to-English or both types in L2 writing tests.

4. Method

4.1 Participants

The participants of study consisted of 60 EFL learners, selected nonrandomly from a larger sample of EFL learners in the Jahad-e-

Daneshgahi Language Center in Shahrekord. The participants, aged 16-20, were native speakers of Persian. The sampling selection was done based on their English proficiency scores on the Nelson English Language Proficiency Test (Section 3000D). Meanwhile, the sample had already taken a placement test developed by the EFL teachers in the language center. Accordingly, they had been placed in advance level courses by the language center. All the participants were studying English as a foreign language in the language center for more than nine months and had already studied *New Interchange Series* (Richards, Hull & Proctor, 2005), where they learned how to write essays in English. Thus, they were all expected to have adequate knowledge about how to write essays in English.

4.2 Instrumentation

For data collection, Nelson English Language Proficiency Test (Section 3000D) and two timed-witting essays were used. The Nelson test, which was used to select the participants, included fifty multiple-choice items to assess the lexical, grammatical and phonological knowledge of the participants. The reliability of the Nelson test as measured by Cronbach's alpha was 0.90. The essays included two topics, not requiring any special knowledge: one on the global warming and the other on the importance of foreign/second language in the world. Care was taken to make the titles of essays as similar as possible in terms of complexity of language, readability and frequency of words. Meanwhile, each essay comprised 350-450 words.

4.3 Procedure

First, the Nelson English Language Proficiency Test was given to a group of 89 EFL learners who had studied English in the institute for more than nine months in the winter semester of 2010. Sixty EFL participants whose proficiency scores were between one standard deviation below and above the group mean score were selected as the main participants of the study. To make sure that they were homogenous,

Leven's test was also conducted on their proficiency scores, which indicated the desirability of results. Second, a comparative 'within-subjects' method was employed to provide comparative evidence from the same group of test takers in two test conditions: writing 2 essays (one with and one without a bilingual dictionary). Third, the participants ($n = 60$) were placed into one of the four groups ($n = 15$), as displayed in Table 1, to control for any differences in performance that might be due to the order or test effects. As Ary, Jacobs and Razavieh (1996) point out, participants in comparative within-subject studies where there are 2 testing conditions can sometimes do better on the second task/test since they may have practice with the first task/test or they do better on the second task/test since they might find it easier. Hence, counterbalancing was used in this study.

Table 1: Order of the test

Group	First Task (Topic 1)	Second Task (Topic 2)
1	Test 1 with dictionary	Test 2 without dictionary
2	Test 1 without dictionary	Test 2 with dictionary
3	Test 2 with dictionary	Test 1 without dictionary
4	Test 2 without dictionary	Test 1 with dictionary

From an original sample of 60 EFL students, 15 participants were allocated to each group. As Table 1 shows, Group 1 wrote two essays: on the first topic (i.e., the global warming) with the use of a dictionary and on the second topic (i.e., the importance of foreign/second language) without the use of a dictionary. That is, the first test was carried out with the use of a dictionary, and the second test was carried out without the use of a dictionary. In Group 2, the order for the dictionary use changed; that is, Group 2 wrote on the first topic (i.e., the global warming) without the use of a dictionary, and on the second topic (i.e., the importance of foreign/second language) with a dictionary. In contrast to Groups 1 and

2, the second topic was given first in Groups 3 and 4; that is, the order for the topic presentation changed. Group 3 did the second test (i.e., writing on the importance of foreign/second language) with the use of a dictionary and the first test (i.e., writing on the global warming) without the use of a dictionary, but Group 4 wrote the second topic with the use of a dictionary, and then the first one without the use of a dictionary; that is, the order for the dictionary use was in reverse in Groups 3 and 4. Therefore, any change in performance could be due to intervention of availability of a bilingual dictionary, not the result of order or test effects. Meanwhile, the participants were asked to indicate the type of dictionary they used, but they were given freedom to choose the dictionary they wanted to use, and even encouraged to use the one with which they had developed a sense of familiarity. Moreover, they had been asked to just underline in their response words looked up in the dictionary to check, for instance, the spelling and grammar or to ascertain the meaning or word choice, but this information was removed from the essays presented to the raters of the essays. Underlining was used to enable quantitative analysis in addressing the second research question. As East (2007) suggests from the result of a pilot study, underlining some words, while writing essays, does not interfere with the writing activity.

Fourth, the essays, written in 50-minute sessions, were transcribed as word processor files and index numbers were given to them. In the copies of files given to the raters of essays, no words looked up in the dictionary had been underlined. Thus, it was difficult for the raters to guess the test condition with accuracy. Meanwhile, the numbers assigned to the essays helped the researchers to identify the essays and conditions in which they were written. Fifth, the tests were scored by two independent raters, both with extensive experience of teaching English at language schools. The raters were not informed which essays had been written in which condition (with and without a dictionary) to reduce researcher sensitivity. The reliability of essay writing tests is heavily influenced by the scoring procedures, including the rating scale to be used and the success with which raters can apply the scale. Therefore, the

raters were given a copy of Jacobs et al.'s (1981) *ESL Composition Profile*, which provided detailed information about the scoring of each criterion in the writing tests. The profile included content, organization, vocabulary, language use, and mechanics criteria. The raters were asked to score each essay, one at a time, according to the criteria, and provide a score for each criterion too. Two weeks later, the two raters were invited again to re-rate just 10 essays to provide some measure of intra-reliability of the scoring. Finally, after discrepancies were discussed by the two raters, the average of two sets of scores from the two raters was reported as the scores for the data analysis.

5. Results

Two sets of independent scores were obtained from two raters for each participant's essays. Coefficients for the inter-rater reliability of the scores given were calculated. The inter-rater reliability coefficient was $r = .92$ $n = 120$, $*p \leq .05$ (two-tailed). Also, intra-rater reliability coefficient for the first and second raters were $r = .95$ and $r = .96$ $n = 10$, $*p \leq .05$ (two-tailed) respectively. Therefore, there was a high level of inter-rater and intra-rater correlation, commensurate with the requirements for the reliable scoring by Hatch and Lazaraton (1991) and Larson-Hall (2010).

To address the first null hypothesis of the study, the final scores were compared across 2 test conditions (with and without a dictionary). As Table 2 displays, the mean of 'with dictionary' scores was a little greater than 'without dictionary' scores, but the standard deviation of 'with dictionary' scores was a little lower. The scores awarded ranged from 51 to 94 and from 48 to 95 in the two test conditions, indicating a wide range of performance in both test conditions. Besides, a comparison of mean scores, as displayed in Table 3, revealed a meaningful difference between scores in the two conditions, $t(59) = 2.67$, $*p \leq .05$ (two-tailed).

Table 2: Descriptive statistics of 'with' and 'without dictionary' writing scores

<i>Variable</i>	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
<i>With dictionary</i>	60	51	94	76.57	13.89
<i>Without dictionary</i>	60	48	95	75.33	14.20

Table 3: Paired sample test between 'with' and without dictionary' writing scores

Variable	t	df	Sig.
With & Without dictionary	2.67	59	.010

For the purpose of the second null hypothesis, the frequency of dictionary use was obtained by the number of 'look-ups' made by each participant. It ranged from 1 to 10, with a mean of 5. Those who had more than 5 look-ups were considered as *high* group, suggesting high frequency of dictionary use and those who had less than 5 look-ups were considered as *low* group in the data analysis, suggesting low frequency of use. Table 4 displays descriptive statistics by frequency of dictionary use and type of bilingual dictionary. As Table 4 demonstrates, the writing mean scores of those who used bilingual dictionaries less frequently were higher ($M = 82.61$) than those who used them more frequently. Besides, the writing mean score of those used English-to-Persian ($M = 87.93$) was higher than the mean scores of those who used other types.

Table 4: Descriptive statistics of 'with dictionary' writing scores by frequency and type of dictionary use

Variable		N	M	SD
Frequency	Low	38	82.61	11.88
	High	22	67.68	11.84
Type of dictionary	Persian-to-English	31	74.32	13.33
	English-to-Persian	14	87.93	9.68
	Both	15	72.87	13.55

To address the second and third null hypotheses, a 2×3 ANOVA was used. The frequency of using a bilingual dictionary and type of bilingual dictionary were considered as independent variables in the analysis.

Table 5: Two-way analysis of variance

Source	Sum of Squares	df	M Square	F	Sig.
Corrected Model	5053	5	102	8.61	.000
Intercept	248417	1	248419	2118	.000
Frequency	1514	1	1514	12.91	.001
Type of Dictionary	1623	2	812	6.92	.002
Type of Dictionary by frequency	184	2	92	.78	.461
Error	6216	53	117		

As Table 5 shows, the frequency of using a bilingual dictionary was significant, $F(1, 53) = 12.91$, $*p \leq .05$, indicating that low or high frequency of bilingual dictionary use made a significant difference in the writing test scores. Also, the type of bilingual dictionary was significant $F(2, 53) = 6.92$, $*p \leq .05$, indicating a significant difference among the scores of the participants who used English-to-Persian, Persian-to-English, or both types of dictionaries.

6. Discussion

The debate over the monolingual dictionary use in testing has recently been shifted to the bilingual dictionary use. Using bilingual dictionaries, according to the results reported in Tables 2 and 3, made a significant difference in test takers' L2 writing test performance. In this study, writing mean scores in the with a dictionary test condition were greater than writing mean scores in the without a dictionary test condition. The EFL participants of this study performed better when they used a Persian-to-English/English-to-Persian dictionary in the L2 writing essays. Therefore, the first null hypothesis of the study is rejected.

If writing tests emphasize outcome which provides us with the manifestation of the knowledge possessed by the writer, as Atwell (1985) states, this knowledge will partly be the result of activating the knowledge of vocabulary, grammar and spelling by dictionaries. In this view, it can be assumed that using bilingual Persian-to-English or English-to-Persian dictionaries activated this knowledge and improved the L2 writing performance. The above results, in contrast to East's (2005, 2008) findings, correspond with Hurman and Tall's (1998; cited in Tall & Hurman, 2002) findings indicating that test scores are significantly higher when bilingual dictionaries are available. At the same time, it supports East's (2005) argument that dictionary can be available as a means of improving the quality of what test takers write, enabling them to be imaginative, creative and more experimental with language, and enhancing their positive experience of the tests.

Messick (1989) suggests two major threats to a test's validity. One is 'construct under-representation', whereby a test fails to include important facets of the construct and the results of a test, thus, might not reveal the test takers' true ability, supposed to have been assessed by the test. The other one is 'construct-irrelevant variance', where the test measures the variables not relevant to the construct in question. This threat allows some test takers to score higher or lower than they would normally score. In this study, the significant differences in the test scores between the two test conditions (i.e., with and without a dictionary) were found. This

would indicate that there might be a threat to the construct validity of L2 writing in the two test conditions. As far as testing is concerned, using bilingual dictionaries in the writing tests might have been a confounding variable since 'with dictionary' scores were found to be significantly higher. In other words, dictionary use might have contributed to the construct-irrelevant variances, hence the increase in the participants' 'with dictionary' mean scores. The other possibility, although subject to much controversy, is that unavailability of bilingual dictionaries might have led to the construct under-representation; that is, the writing construct was under-represented in the without a dictionary test condition. As it was observed, the 'without dictionary' writing scores were significantly lower than the 'with dictionary' writing scores, given the steps taken in this study to nullify the effect of extraneous factors such as order and test effects. However, it should be taken into cognizance that construct validity, in a broader sense, as Messick (1989) states, "binds other social consequences of testing to the evidential basis of test interpretation and use" (p. 21). Thus, other factors should be considered since their impact on the testing procedure may affect the construct validity of the tests. In other words, there might be good reasons, other than test scores in assessment, to make decision to allow or ban use of dictionary in the writing tasks.

According to the results in Table 5, frequency of dictionary use had a significant effect on writing test scores. Hence, the second null hypothesis of the study is rejected. However, the effect of the frequency variable is not always positive. The high frequency of bilingual dictionary use does not ensure the better performance since those who used bilingual dictionaries less than 5 times showed a better performance. They might have spent less time seeking words in the dictionaries and, instead, tried to use their own knowledge to convey and organize their thoughts and meanings. Perhaps, using bilingual dictionaries so frequently took too long in the tests. For those who used it more than 5 times, the use of dictionary in the test might have interfered with the tests' measurement properties; As East (2005) argues, it might have

distracted them from a real knowledge. The above issue suggests that the more able participants use bilingual dictionaries to improve their performance, but the most able ones use them less frequently. The findings of this study are in agreement with those of Atkins and Varantola (1998) who found out that lower ability language learners would make more use of the dictionary than higher ability learners. Perhaps, one should agree with East (2007), who has reported that frequency variable fails to predict writing performance. Thus, too much reliance on bilingual dictionaries can reduce learner autonomy in L2 writing classes.

Finally, the type of dictionary was found to be a significant variable in this study. Thus, the last null hypothesis of this study is rejected too. The test takers who had used English-to-Persian dictionaries received the highest scores, and the test takers who had used both types of bilingual dictionaries (English-to-Persian and Persian-to-English) received the lowest scores. It can be assumed that L2-to-L1 dictionaries are more useful in L2 writing tasks. Perhaps, those who use both types of dictionaries feel less confident and waste more time with them to check mechanics, words and grammar. This might interfere with transfer of semantic knowledge in the process of writing. It is also likely that those who prefer one type, such as English-to-Persian, access better strategies in using dictionaries in L2 writing than those who use both types.

7. Conclusion

There are conflicting views about whether the use of dictionary—monolingual or bilingual—is masking the measurement of the test takers' witting ability or not. There might be good reasons for L2 teachers to either exclude or include bilingual in L2 writing tests. The study reported here on the availability of bilingual dictionaries has revealed that performance, as reflected in the timed-writing essay mean scores, improves when Persian-to-English and/or English-to-Persian dictionaries are used. This finding serves to highlight the positive role of bilingual dictionaries in the writing performance. They also enable us to consider

arguments for including the dictionary in the classroom writing activities. However, the frequent use of bilingual dictionaries in the test conditions might not benefit EFL learners as it might lead to confusion or interfere with the other components of writing construct, such as organization and content. The type of bilingual dictionary was also found to be a significant variable in this study. English-to-Persian dictionaries, as reflected in the higher mean scores, seemed to be more useful than Persian-to-English- ones in the writing performance. Whether this can be due to the type of information provided in the bilingual dictionaries, quality, strategies or just perceptions of test takers is not still quite clear.

8. Implications, Limitations and Suggestions

Based on the findings of this study, L2 teachers in Iran can allow learners, particularly the less advanced ones, to make use of bilingual dictionaries to demonstrate their writing skill in L2 writing classrooms. As East (2005) pointed out, using dictionaries in assessment has been the subject of debate. If, however, focus is placed on helping L2 teachers to use assessment, as part of teaching and learning in classrooms, in ways that will raise their learners' achievement, dictionaries may have a valid role to play. Therefore, it is suggested that Iranian EFL learners, in particular, make use of Persian-to-English or English-to-Persian dictionaries as a supportive tool in essay writing classes. Another implication of this study is that product approaches to writing should not be totally abandoned. Rather, they should be complementary to the process approaches; bilingual dictionaries are useful in the process as well as the product of L2 writing. EFL learners should be taught not only heuristic devices to focus on meaning, but also heuristic devices to focus on linguistic features. However, the findings obtained in this study imply that high frequency of dictionary use, though important, does not compensate for the lack of L2 writing knowledge.

One limitation of this study is that it did not probe why and how the participants preferred or used their bilingual dictionaries. Perhaps a post-interview or retrospective data collection procedure would have revealed

more information about the strategies of dictionary use and usefulness of the type of bilingual dictionaries. At the very least, the findings of this study indicate that the English-to-Persian dictionaries proved more useful than Persian-to-English in L2 writing tests. Another study is required to probe the psychological benefits and perceptions of L2 learners about types of bilingual dictionaries. The complex picture presented here is sufficient to suggest further research on the above issue.

References

- Albus, D., Thurlow, M., Liu, K., & Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research*, 98(4), 245-254.
- Ary, D., Jacobs, L. C., & Razavieh, A. (1996). *Introduction to research education*. Fortworth, USA: Harcourt Brace College Publishers.
- Atkins, B. T. S., & Varantola, K. (1998). Monitoring dictionary use. In B. T. S. Atkins (Ed.), *Using dictionaries: Studies of dictionary use by language learners* (pp. 83-122). Germany: Max Niemeyer Verlag.
- Atwell, N. (1985). How we learned to write. *Learning*, 13, 51-53.
- Bachman, L. F. (1990). *Fundamental considerations in language learning*. Oxford: Oxford University Press.
- Bensoussan, M., Sim, D., & Weiss, R. (1984). The effect of dictionary usage on EFL test performances compared with student and teacher attitudes and expectations. *Reading in a Foreign Language*, 2(2), 262-276.
- Bishop, G. (1998). Research into the use being made of bilingual dictionaries by language learners. *Language Learning Journal*, 18(1), 3-8.
- Chambers, G. (1999). Using dictionaries in the GCSE examination of modern foreign languages-pupils' perceptions. *Studies in Modern Language Education*, 7(1), 68-85.
- Chastain, K. (1988). *Developing second language skills: Theory and practice*. New York: Harcourt Collage Publishers.

- Conner, V. (1987). Argumentative patterns in students essays: Cross-cultural differences. In U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text reading* (pp. 57-72). Massachusetts: Anderson Wesley.
- Cumming, A. (2002). Assessing L2 writing: Alternative constructs and ethical dilemmas, *Assessing Writing*, 8(2), 73-83.
- East, M. (2005, September). *Autonomy in an assessment context: Allowing dictionaries in writing tests*. Paper presented at the meeting of Independent Learning Association Conference Inaugural on the Supporting Independent English Language Learning in the 21st Century, Auckland.
- East, M. (2006). The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: A quantitative analysis. *Assessing Writing*, 11(3), 179-197.
- East, M. (2007). Bilingual dictionaries in tests of L2 writing proficiency: Do they make a difference? *Language Testing*, 24(3), 331-353.
- East, M. (2008). Dictionary use in foreign language writing exams: Impact and implications. *International Journal of Lexicography*, 22(3), 344-348.
- Elola, I. (2008). Dictionary use and vocabulary choices in L2 writing. *Estudios de lingÜística inglesa aplicada*, 8, 63-89.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139-155.
- Hartmann, R. R. K. (1990). Sociology of the dictionary user: Hypotheses and empirical studies. *An International Handbook*, 8(1), 23-43.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston: Heinle and Heinle.
- Hayati, A. M. (2005). A comparative study of using bilingual and monolingual dictionaries in reading comprehension of intermediate EFL students. *The Reading Matrix*, 5(2), 331-353.

- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university level. *The Modern Language Journal*, 80(3), 309-326.
- Hillocks, G. (1986). Synthesis of research on teaching writing. *Educational Relationship*, 44(8), 71-82.
- Horsfall, P. (1997). Dictionary skills in MFL 11-16. *Language Learning Journal*, 15(1), 3-9.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley: Newbury House.
- Kern, R., & Schultz, J. (1992). The effects of composition instruction on intermediate level French students' writing performance: Some preliminary findings. *Modern Language Journal*, 76(1), 1-13.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Martin, R. (2002). One benefit of monolingual dictionaries in the writing classroom. *TESL Journal*, 12(1), 8-20.
- Messick, S. (1989). Validity. In R. R. Linn (Ed), *Educational measurement* (pp. 13-103). New York. Macmillan.
- Raimes, A. (1990). The TOEFL test of written English: Causes for concern. *TESOL Quarterly*, 24(3), 427-442.
- Richards, J. C., Hull, J., & Proctor, S. (2005). *New interchange English for international communication*. Cambridge: Cambridge University Press.
- Rivera, C., & Stansfield, C. W. (1998). *Leveling the playing field for English language learners: Increasing participation in state and local assessments through accommodations*. Retrieved August 4, 2006, from http://ceee.gwu.edu/standadads_assessments/researchLEP-accomodrinto.html
- Tall, G., & Hurman, J. (2002). Using dictionaries in modern languages GCSE examinations. *Educational Review*, 54, 205-217.
- Thompson, G. (1987). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, 22(1), 58-78.

- Widdowson, H. (1983). *Learning purpose and language use*. Oxford: Oxford University Press.
- Zamel, V. (1983). The composing processes of advanced ESL students: Six cases studies. *TESOL Quarterly*, 17(2), 165-187.
- Zamel, V. (1987). Recent research on writing pedagogy. *TESOL Quarterly*, 21(4), 697-715.

