Effects of Different Response Types on Iranian EFL Test Takers' Performance

Mohammad Hassan Chehrazad
PhD Candidate, University of Tabriz
chehrazad88@ms.tabrizu.ac.ir

Parviz Ajideh
Professor, University of Tabriz
parvizaj@gmail.com

Abstract

Test method facet is one of the factors which can have an influence on the test takers' performance. The purpose of the current study was to investigate the effects of two different response types, multiple-choice cloze and multiple-choice test, on the pre-intermediate and intermediate test takers' reading comprehension performance. To this end, 40 pre-intermediate and intermediate learners participated in the study. To counterbalance the practice effect, the participants of the study at both pre-intermediate and intermediate levels were randomly assigned into two groups. For each level, there were two types of tests and two test administrations. Within each level, the two groups took these two tests in a different order. For all participants of the study there were two scores based on multiple- choice cloze and multiple-choice tests. Two Paired sample t tests were separately run to compare the pre-intermediate and intermediate test takers' performance on multiple-choice cloze and multiple-choice tests. Two independent samples t tests were run to compare the pre-intermediate and intermediate test takers on these two tests. There was no significant difference between the pre-intermediate test takers' performance on these two types of tests. There was a significant difference between intermediate test takers' performance on the tests. There was no significant difference between the pre-intermediate and intermediate test takers' performance on the multiple-choice cloze tests. The intermediate test takers outperformed the pre-intermediate test takers only on the multiple-choice tests. Based on these results it can be concluded that test methods can have an influence on the test takers' performance, especially at higher proficiency levels and learners may need to have reached a certain level of proficiency to be able to understand the text.

Keywords: Response Types, Multiple-Choice Cloze, Multiple-Choice Test

Received: January 2012; Accepted: December 2012

1. Introduction

According to Bachman (1990):

> Language test performance is affected by different factors, and an understanding of these factors and how they affect test scores is fundamental to the development and use of language tests. Test performance can be influenced by communicative language ability, personal attributes which are not part of language abilities we are interested in, random factors which are unpredictable and temporary, and test method facets. (1990, p. 164)

Of these different factors, according to Bachman (1990) and Bachman and Palmer (1996), which can affect test takers' performance, test method is one of the most important factors that have attracted language measurement specialists and teachers' attention, and directed their focus to its importance and to its effects on both the test takers' performance and the quality of the obtained information.

Language tests are important because of the following reasons. One reason that they are important, according to Bachman (1990), is related to their potential to be used as means for controlling the context in which language performance takes place. The characteristics of test methods can be seen as restricted and controlled versions of these contextual features that determine the nature of performance that is expected for a given test or test task. Based on the role of the contextual features in language use in general, it is not surprising to find that aspects of the test method, which provide much of the context of language tests, affect performance on language tests.

Another reason is related to the quality of tests. According to Shohamy (1984), a good test is a test in which the method has little effect on the ability being tested. That is, if test takers' performance on a test is the result of the

ability being measured rather than the testing method, that test is considered to be a good testing tool.

The next reason, based on Bachman (1990), is based on the characteristics or facets of test methods which constitute the how of language testing, and are of particular importance for designing, developing, and using language tests. Furthermore, teachers, within their teaching process, need to obtain information about the students to assess their achievement and improve their teaching by applying the results. To use language tests for these purposes and to make decisions, the quality of the information upon which the decisions are based must be reliable and relevant.

The most significant reason, as appreciated by Bachman (1990) and Bachman and Palmer (1996), is related to the role of test methods in learners' performance on a given test. They influence different test takers' performance differently and systematically. Test method facets are systematic to the extent that they are uniform from one test administration to the next. That is, if the input format facet is multiple-choice, this will not vary, whether the test is given in the morning or afternoon. Considering their different effects on test takers' performance, some test takers, for example, may perform better in the context of an oral interview than they would be sitting in a language laboratory and speaking into a microphone in response to statements presented through a pair of earphones.

Because of the importance of the effects of test method facets on test performance, Bachman (1990) has developed a framework for delineating the specific features, or facets of test method. The five major categories of test method facet are:

1. The testing environment
2. The test rubric

3. The nature of the input
4. The nature of the expected response
5. The relationship between the input and the response.

1.1. Background (Empirical Studies)

Shohamy (1983) studied the effects of different aspects of test method facets on test takers' performance and have demonstrated that the methods we use to measure language ability influence performance on language tests and Vygotsky (1969) found a relationship between the language of test instructions and test-takers' performance. Bachman and Palmer(1981a) also found that scores from self-ratings loaded consistently more highly on method factors than on specific trait or ability factors, and that translation and interview measures of reading loaded more heavily on method than on trait factors. In addition to, Bachman and Palmer (1982a) also found that scores from both self-ratings and oral interviews consistently loaded more heavily on test method factors than on specific trait factors. One of such factors, test method facets, is the influence of "test format". Whether test constructors use "multiple-choice", "true-false", "open-ended" or other testing formats in their tests, may influence the test takers' performance (e.g., Alderson, 2000; Bachman & Palmer, 1996; Buck, 2001).

Kintsch and Yarbrough (1982), for example, investigated the effects of two test formats, open-ended questions and cloze tests on reading comprehension test performance. They suggested that open ended questions can measure the reader's comprehension of main ideas of the text, whereas cloze tests will touch only upon local understanding and will not reflect the readers' overall comprehension. Moreover, Shohamy (1984) examined the effect of various test methods, namely multiple-choice and open ended questions measuring reading

comprehension. Results of her study revealed that each of the test facets produced different degrees of difficulty for subjects and that each of the variables, method, text, and language, had a significant effect on students' performance on the test of reading comprehension especially with low level students suggesting further support for the role of language proficiency on test takers' performance on reading comprehension test.

Furthermore, Buck (1990, 1991) examined the effects of item stem preview on test takers' performance through a comparison between the mean scores of the groups who previewed item stems and those who did not. Interestingly, neither study found any significant effect for item stem preview on test taker performance or item difficulty. Kobayashi (2002) also addressed the effects of test method facets such as text organization and response format. He found that text organization and test format had a significant impact on the Japanese university students' performance of reading comprehension tests, and with an interaction between the two variables. His study further revealed that more proficient learners performed better in summary writing and open-ended questions with clearly organized texts.

In addition, Jafarpur (2003) explored the relative effect of test developer on the performance of test takers using multiple choice reading comprehension tests that had no specifications. He concluded that there may be a facet of test constructor. Lumley and O'Sullivan (2005) also investigated the role of interaction of variables such as the task topic and the gender of the person presenting the items and the gender of test takers on a tape-mediated test of speaking ability and found that the effects of interactions were small, the role of the gender of the interlocutor was limited, and the effect of the task type was slightly more significant.

Moreover, In'nami and Koizumi (2009) conducted a meta-analysis on the impacts of test facets, namely multiple-choice and open-ended questions on performance on L1 reading, L2 reading, and L2 listening. In general, they found multiple-choice formats easier than open-ended questions in both L1 reading and L2 listening while no impact of test formats was found in L2 reading.

Despite these studies, one of the major limitations of the research to date is that there is no information available on the investigation of the effects of the selected response types, as an aspect of test method facet, on the Iranian EFL test takers' test performance.

2. Methodology

2.1. Purpose of the Study

The main purpose of the current study was to investigate the different effects of two selected response types, multiple-choice cloze and multiple-choice reading comprehension tests, on the pre-intermediate and intermediate test takers' performance

2.2. Research Questions

Research Question 1(RQ1): Is there any difference between pre-intermediate test takers' performance on multiple-choice cloze and multiple-choice reading comprehension tests?

Research Question 2(RQ2): Is there any significant difference between intermediate level test takers' performance on multiple-choice cloze and multiple-choice reading comprehension tests?

Research Question 3(RQ3): Is there any difference between pre-intermediate and intermediate level test takers' performance on multiple-choice reading comprehension tests?

Research Question 4(RQ4): Is there any difference between pre-intermediate and intermediate level test takers' performance on multiple-choice cloze tests?

2.3. Participants

This study was conducted in an English institute in Tabriz, Iran. The number of learners who participated in the study was 40. They were from Tabriz and bilingual speakers of both Persian and Azeri. All of the participants were male and between 14-20 years old. Half of the participants were at the pre-intermediate level and half of them were at the intermediate level. In this system, the learners are called pre-intermediate and intermediate after studying for 6 terms, 1 year, and 12 terms, 2 years respectively. In addition, a general English test, KET (Key English Test), was used to systematically establish participants' homogeneity. Based on the result of this test, most of the participants of the pre-intermediate and intermediate levels' scores were respectively between 25-35 and 40-55. There were few participants, at both levels, whose scores were either higher or lower than these scores and, consequently, they weren't included in the study

2.4. Materials

2.4.1. Reading Comprehension Texts

Since one of the purposes of the study was to compare pre-intermediate and intermediate test takers' reading comprehension performance on different response format, it was essential to include texts which were at different levels

of difficulty and which were appropriate for their proficiency levels. To this end, to control the effects of the level bias, two different texts, for each of the levels, resulting in a total of 4 texts, were selected. The chosen topics for the pre-intermediate level were 'A Postman in India' and 'Travelling in India'. The chosen topics for the intermediate level were 'The Reasons to Talk about Weather' and 'The Power of Makeup'. These texts and topics were from their textbooks, and were based on their proficiency levels. Although these texts were from their textbooks, they hadn't been covered in the class at the time of the test administrations. That is, the first test administration was the participants' first exposure to these texts.

2.4.2. Tests

For each of the four reading comprehension texts, two texts for the pre-intermediate level participants and two texts for the intermediate level participants, two types of tests were developed. One of these tests was a four option multiple- choice test which was based on the comprehension of the texts. The other test was a standard multiple choice cloze test which was developed by deleting every 7th word of the text. It was a four option multiple-choice cloze test. To develop the test items and their distractors, for both pre-intermediate and intermediate levels the tests were given to other test takers who were at the same proficiency level as the main participants of the study. Their wrongly chosen answers were used as the distractors of the test items. The developed tests, eight tests, for all texts and for both pre-intermediate and intermediate level participants are illustrated in Table 1:

Table1. Tests Developed for Pre-Intermediate and Intermediate Participants

Proficiency level	Topic	Type of test
Pre-intermediate	A postman in India	Multiple-choice test
Pre-intermediate	A postman in India	Multiple-choice cloze test
Pre-intermediate	Travelling in India	Multiple-choice test
Pre-intermediate	Travelling in India	Multiple-choice cloze test
Intermediate	The reasons to talk about weather	Multiple-choice test
Intermediate	The reasons to talk about weather	Multiple-choice cloze test
Intermediate	The power of makeup	Multiple-choice test
Intermediate	The power of makeup	Multiple-choice cloze test

As it is clear from the table, for all texts at both levels, two types of tests were developed. In order to counterbalance the practice effect, the participants at these two levels, pre-intermediate and intermediate, were divided into half, two groups, and were given the tests in the order and design which is shown in Table 2.

Table 2. The Order and Design of the Given Tests

Proficiency level	**First test administration**	**Second test administration**
Pre-intermediate Group 1	A postman in India/ 8 items & multiple-choice test	A postman in India/ 25 items & multiple-choice cloze test
	Travelling in India/ 25 items & multiple-choice cloze test	Travelling in India/ 8 items & multiple-choice test
Pre-intermediate Group 2	A postman in India/ 25 items & multiple-choice cloze test	A postman in India/ 8 items & multiple-choice test
	Travelling in India/ 8 items & multiple-choice test	Travelling in India/ 25 items & multiple-choice cloze test
Intermediate Group 1	The reasons to talk about weather/ 8 items & multiple-choice test	The reasons to talk about weather/ 25 items & multiple-choice cloze test
	The power of makeup/ 25 items & multiple-choice cloze test	The power of makeup/ 8 items & multiple-choice test
Intermediate Group2	The reasons to talk about weather/ 25 items & multiple-choice cloze test	The reasons to talk about weather/ 8 items & multiple-choice test
	The power of makeup/ 8 items & multiple-choice test	The power of makeup/ 25 items & multiple-choice cloze test

The logic behind including eight test items for multiple-choice cloze reading comprehension tests was following the standard procedure of including this number of items in most of the standard tests' reading comprehension section. As it is depicted in the table, all pre-intermediate and intermediate level participants of the study took two types of tests. One of these tests was a multiple-choice cloze test and the other one was a multiple-choice test.

2.5. Procedure

After the participants and the materials were chosen, the procedure commenced. Although the participants' proficiency level was based on the institute's system, and accordingly they were either at pre-intermediate or intermediate levels, in order to ensure of their proficiency level, a general English test, KET (Key English Test), was used. Based on the result of this test, most of the participants of the pre-intermediate and intermediate levels' scores were between 25-35 and 40-55. There were few participants, at both levels, whose scores were either higher or lower than these scores and, consequently, they weren't included in the study.

One session after the administration of KET, the previously prepared texts and tests were given to the participants. The second test administration was one month after the first administration. It is essential to mention that the difference between these two test administrations was related to the different types of tests, based on the same texts for each level, which were given to the two different groups of both pre-intermediate and intermediate levels.

All scores of all participants, for multiple-choice cloze and multiple-choice tests, at both pre-intermediate and intermediate levels, were summed. Consequently, for both pre-intermediate and intermediate level participants, there were two scores which were based on the similar texts. One of the scores

was based on the multiple-choice cloze test. The other score was based on the multiple-choice test. That is, every participant had two scores which were based on the tests of the same texts. One of these scores was on the multiple-choice cloze test and the other one was based on the multiple-choice test of the same texts.

2.6. Statistical Analysis

After the final scores of all pre-intermediate and intermediate participants, which were based on two multiple-choice cloze and multiple-choice tests of the same texts, were calculated, they were analyzed using SPSS. For both tests at pre-intermediate and intermediate levels descriptive statistics were calculated. In order to test the first research question, the existence or nonexistence of the significant difference between the pre-intermediate test takers' performance on the multiple-choice cloze and multiple-choice tests, a paired sample t test was used. Another paired sample t test was used to the test the second research question which was the presence or absence of a significant difference between the intermediate test takers' performance on the multiple-choice cloze and multiple-choice tests. In order to test the third research question, the presence or absence of the significant differences between the pre-intermediate and intermediate level test takers' performance on multiple-choice tests, an independent samples t test was used. Another independent samples t test was used to test the fourth research question, the presence or absence of the significant differences between the pre-intermediate and intermediate level test takers' performance on multiple-choice cloze tests.

2.7. Results

The results of the descriptive statistics for the pre-intermediate test takers' performance on two types of tests, multiple-choice cloze and multiple-choice reading comprehension tests, are given in Table 3:

Table 3. Pre-Intermediate Test Takers' Performance on Two Types of Tests

Level	Test	Mean	N	Std. Deviation	Std. Error Mean
Pre-intermediate	Multiple-choice cloze	23.6433	21	5.89780	1.28700
	Multiple-choice test	22.9724	21	6.43959	1.40523

As it is clear from the table, the pre-intermediate test takers' mean on multiple-choice cloze test was more than their mean on multiple-choice test. In addition, their standard deviation on multiple-choice cloze was less than their standard deviation on multiple-choice test. Based on these differences, their performance on multiple-choice cloze was descriptively better than it was on multiple-choice test.

The results of the paired samples test for the pre-intermediate test takers' performance on two types of tests, multiple-choice cloze and multiple-choice reading comprehension tests, are given in Table 4:

Table 4. Paired Samples Test of Pre-Intermediate Test Takers' Performance on Two Types of Tests

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Multiple-choice cloze & Multiple-choice test	.67095	5.00173	1.09147	-1.60581	2.94771	.615	20	.546

Since p-value=.546>α=0.05, the pre-intermediate test takers' different performance on these two types of tests, multiple-choice cloze and multiple-choice reading comprehension tests, is not statistically significant.

The results of the descriptive statistics for the intermediate test takers' performance on two types of tests, multiple-choice cloze and multiple-choice reading comprehension tests, are given in Table 5:

Table 5. Intermediate Test Takers' Performance on Two Types of Tests

Level	Test	Mean	N	Std. Deviation	Std. Error Mean
Intermediate	Multiple choice cloze	26.1635	20	4.54582	1.01648
	Multiple choice test	29.6030	20	6.92044	1.54746

As it is clear from the table, in contrast to the pre-intermediate test takers, the intermediate test takers' mean on multiple-choice test was better than their mean on multiple-choice cloze. In addition, the intermediate test takers' standard deviation on multiple-choice cloze was less than their standard deviation on multiple-choice test.

The results of the paired samples test for the intermediate test takers' performance on two types of tests, multiple-choice cloze and multiple-choice reading comprehension tests, are given in Table 6:

Table 6. Paired Samples Test of Intermediate Test Takers' Performance on Two Types of Tests

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Multiple-choice cloze & Multiple-choice test	-3.43950	6.53637	1.46158	-6.49862	-.38038	-2.353	19	.030*

Since p-value=.030<α=0.05, the intermediate test takers' performance on these two types of tests, multiple-choice cloze and multiple-choice reading comprehension tests are significantly different.

The results of the descriptive statistics for the comparison of pre intermediate and intermediate test takers' performance on multiple-choice reading comprehension tests are depicted in Table 7:

Table 7. Pre-Intermediate and Intermediate Test Takers' Performance on Multiple-Choice Tests

Level	Test	N	Mean	Std. Deviation	Std. Error Mean
Pre-intermediate	Multiple-choice test	21	22.9724	6.43959	1.40523
Intermediate		20	29.6030	6.92044	1.54746

According to the table, intermediate test takers descriptively outperformed pre-intermediate test takers on multiple-choice reading comprehension tests.

The results of the first independent samples tests for the pre-intermediate and intermediate test takers' performance on multiple-choice reading comprehension tests are shown in Table 8:

Table 8. Independent Samples Test for Pre-Intermediate and Intermediate Test Takers' Performance on Multiple-Choice Tests

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.394	.534	-3.178	39	.003	-6.63062	2.08653	-10.85103	-2.41021
Equal variances not assumed			-3.172	38.430	.003	-6.63062	2.09029	-10.86063	-2.40060

Since p-value=.003<α=0.05, the pre-intermediate and intermediate test takers' performance on multiple-choice reading comprehension tests are significantly different.

The results of the descriptive statistics for the comparison of pre-intermediate and intermediate test takers' performance on multiple-choice cloze tests are shown in Table 9:

Table 9. Pre-Intermediate and Intermediate Test takers' Performance on Multiple-Choice Cloze Tests

Test	Level	N	Mean	Std. Deviation	Std. Error Mean
Multiple-choice cloze tests	Pre-intermediate	21	23.6433	5.89780	1.28700
	Intermediate	20	26.1635	4.54582	1.01648

Based on this table, the intermediate test takers' mean on multiple-choice cloze test was better than pre-intermediate test takers' mean on multiple-choice test. In addition, their standard deviation on multiple-choice cloze was less than intermediate test takers' standard deviation on multiple-choice test. Based on these descriptive statistics, intermediate test takers outperformed pre-intermediate test takers.

The results of the second independent samples tests for the pre-intermediate and intermediate test takers' performance on multiple-choice cloze tests are shown in Table 10:

Table 10. Independent Samples Test for Pre-Intermediate and Intermediate Test Takers' Performance on Multiple-Choice Cloze Tests

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	1.632	.209	-1.527	39	.135	-2.52017	1.65048	-5.85858	.81825
Equal variances not assumed			-1.537	37.411	.133	-2.52017	1.64000	-5.84190	.80156

Since p-value=.135>α=0.05, the difference between pre-intermediate and intermediate test takers' performance on multiple-choice cloze tests are not statistically significant.

3. Discussion and Conclusion

In spite of the fact that the pre-intermediate test takers' mean on multiple-choice cloze tests was a little more than their mean on multiple-choice test, this outperformance was not statistically significant. Based on this finding, the first null hypothesis (H01), lack of a significant difference between pre-intermediate level test takers' performance on multiple-choice cloze and multiple-choice reading comprehension tests, was accepted.

In addition, the intermediate test takers' mean on multiple-choice tests was better than their mean on multiple-choice cloze tests. This outperformance is statistically significant. Therefore, the second null hypothesis (H02), lack of a

significant difference between intermediate test takers' performance on multiple-choice cloze and multiple-choice reading comprehension tests, was rejected.

Intermediate test takers outperformed pre-intermediate test takers on multiple-choice reading comprehension tests and this outperformance was statistically significant. Hence, the third null hypothesis (H03), lack of a significant difference between pre-intermediate and intermediate test takers' performance on multiple choice tests, was rejected.

Although the intermediate test takers' mean on multiple-choice cloze test was more than pre-intermediate test takers' mean on multiple-choice cloze test, this difference was not statistically significant. Consequently, the fourth null hypothesis (H04), lack of a significant difference between pre-intermediate and intermediate test takers' performance on multiple-choice cloze tests, was accepted.

Considering these results, test type, whether it is a multiple-choice cloze or a multiple-choice test, does not seem to make a significant difference if the learners' level of language proficiency is pre-intermediate. It can be due to the pre-intermediate test takers' proficiency level which is not high enough to result in their successful performance on reading comprehension tests. It can also be because of their unfamiliarity with the test taking strategies and ways of dealing with reading comprehension texts and tests.

Furthermore, types of tests of reading comprehension make a significant difference if the learners' level of language proficiency is intermediate. That is, intermediate test takers perform better when they are tested by multiple-choice reading comprehension tests than when they are tested by multiple-choice cloze tests. It can be due to their more familiarity with multiple-choice tests, usually developed as a result of having the previous experience of taking these

kinds of reading comprehension tests. It can also be due to their unfamiliarity or less familiarity with multiple-choice cloze tests.

Regarding the interaction of multiple-choice cloze test and language proficiency, there was no interaction and it made no difference whether the level of proficiency was pre-intermediate or intermediate and the test was a multiple-choice cloze test. That is, the pre-intermediate and intermediate test takers' performance on multiple-choice cloze tests was not significantly different. This lack of difference between pre-intermediate and intermediate test takers' performance on multiple-choice cloze tests can be similarly explained by their unfamiliarity with multiple-choice cloze tests. It can be suggested that in spite of the intermediate test takers' higher proficiency than that of the pre-intermediate test takers, it does have no effect on their ability to take multiple-choice cloze tests.

Unlike the lack of the interaction between level of proficiency and multiple-choice cloze tests, there was an interaction between level of proficiency and multiple-choice tests. In other words, when the test was a multiple-choice test, the level of language proficiency affected the test takers' performance. It can be due to the intermediate test takers' more familiarity with multiple-choice reading comprehension tests than the pre-intermediate test takers. It can be suggested that as the learners' level of proficiency develops from pre-intermediate to intermediate, their ability to take the multiple-choice reading comprehension tests develops.

Considering these results and with respect to their comparison with the previous studies, particularly based on the effects of test methods on test performance, they are in line with several studies (Bachman and Palmer 1981a, 1982a; Kobayashi, 2002; Shohamy, 1983) which demonstrated that the methods we use to measure language ability influence performance on language tests.

The results of this study also supports In'nami and Koizumi's (2009) study in which they found multiple-choice formats easier than open-ended ones.

Regarding the interaction of the proficiency and test method facets, the results are in line with Shohamy's (1984) study which supported of the role of language proficiency on test takers' performance on reading comprehension test. They can also provide further support to the concept of linguistic threshold which was previously supported by Claphman (1996) and Ridgway (1997), according to which learners may need to have reached a certain level of proficiency to be able to understand the text.

Regarding the implications of these findings, both test type and proficiency level, and their interaction, are important factors which should be taken into consideration in the development and use of language tests. It can be suggested that when the test takers are at lower levels of proficiency, pre-intermediate, the type of the given test does not make any significant difference and has no significant effect on the test takers' performance. As the level of proficiency gets higher, the effects of the test method facet, particularly the type of expected response, becomes more evident and significantly affects the test takers' performance. Consequently, in the development and use of language tests, the test takers' proficiency level, their familiarity with the test types, and practicality, regarding the preparation and administration, should be carefully taken into account.

Considering the limitations of the current study, it should be mentioned that the current study has investigated the test performance of a small sample of Iranian EFL test takers. All of the participants were male. They were at either pre-intermediate or intermediate levels. There can be another study comparing the performance of male and female test takers on these kinds of

tests. It would be interesting to study whether the findings will apply to learners of higher language proficiency or even native speakers.

In addition, the focus of the current study was on two different response types, multiple-choice cloze and multiple choice tests. Therefore, more research needs to be conducted to study other response types and their effects on the test takers' performance. In addition to these areas, it would be interesting to investigate other aspects of Bachman's model and their effects on the test takers' performance. Further research also needs to be done to study the effects of teaching test taking strategies, both multiple-choice and multiple-choice cloze tests, on the test takers' performance.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Bachman, L. (1990). *Fundamental considerations in language testing.* London: OUP.

Bachman, L. F., & Palmer, A. S. (1981).The construct validation of the FSI oral interview. *Language Learning, 31*(1), 67-86.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some component of communicative proficiency. *TESOL Quarterly, 16*(4), 449-465.

Bachman, L. F., & Palmer, A. S. (1996).*Language testing in practice*. Oxford: Oxford University Press.

Buck, G. (1990). *Testing second language listening comprehension*. Unpublished doctoral dissertation, University of Lancaster.

Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing, 8*(1), 67-91.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Clapham, C. (1996).*The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing, 26*(2), 219-244.

Jafarpur, A. (2003). Is the test constructor a facet? *Language Testing, 20*(1), 57-87.

Kintsch, W. & Yarbrough, J. C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology 74*, 828–834.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19*(2), 193-220.

Lumley, T., & O'Sullivan, B. (2005). The effect of test-taker gender, audience and topic on task performance in tape-mediated assessment of speaking. *Language Testing, 22*(4), 415-437.

Ridgway, T. (1997). Thresholds of the background knowledge effect in foreign language reading. *Reading in a Foreign Language, 11*(1), 151–168.

Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedure. *Language Learning, 33*(4), 527-540.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing, 1*(2), 147-170.

Vygotsky, L. S. (1969). *Thought and language.* Cambridge, MA: MIT Press.