Effect of item variety on the performance of EFL learners' reading comprehension

Bahardoust Maryam, Ghaemi Farid

Corresponding Author email: shadib49571@yahoo.com

ABSTRACT: Reading comprehension can be considered as a multifaceted skill that involves understanding a text, processing information, and interpreting the author's intention. Moreover, item variety has been regarded as one of the most important factors affecting the testee's performance on reading comprehension. The research literature reports that varying test items may differentially influence the performance of EFL learners. The present study's main aim is to investigate the effect of item variety on students' L2 reading comprehension test performance. To this end, 30 EFL learners from one language institute in Kashan, Iran were asked to respond the questions of three reading passages with a general topic of communication. Item varieties following each passage included true/false, multiple-choice, and gap filling questions, respectively. The results indicated that participants' performances were affected by the item variety. It was also found that the difficulty level of test items from the most difficult to the easiest were multiple-choice, true/false, and gap-filling questions, respectively.

Keywords: Reading comprehension; Item variety; Testee's performance; Multiple-choice; True-false; Gap-filling questions

INTRODUCTION

Reading skill can be defined as a receptive decoding language process (Chastain, 1988). While Li and Wilhelm (2008) maintained that reading is a receptive, active, and dynamic process since the readers make attempt to connect the ideas in the text. Also, it is a source of enjoyment for the readers. Reading can be regarded as a basic skill that readers need to process in order to be successful in learning and communication. In other words, reading ability provides reasonable and considerable opportunities for effective communication. Competent readers have sufficient capability to analyse and process concerned text and thereby get the main idea and writer's intention. In order to succeed, second or foreign readers should have high levels of reading proficiency. Therefore, reading effectively is considered as one of the most critical issues in second or foreign language (Carrell, 1989). "A lack of understanding written text can limit possibilities for social and academic success" (Falke, 2008).

McKee (2012) argued that reading in and of itself is insufficient. Despite this fact that reading is considered as an essential part of a reader's personal and educational development, it is the concept of comprehension that may have even much more importance. Reading comprehension is an active, complex and multifaceted skill. However, different scholars and researchers have offered various definitions on the concept of reading comprehension. For example, Falke (2008) claimed that albeit it is generally admitted that reading comprehension is a basic and essential skill for educational, occupational, and social achievement, scholars have not already arrived at a full agreement on what reading comprehension means. For example, Veeravagu, et al (2010) defined reading comprehension as "a thinking process by which a reader selects facts, information, or ideas from printed materials; determines the meanings the author intended to transmit; decide how they relate to previous knowledge; and judge their appropriateness and worth for meeting the learner's own objectives"(p.206). According to Grab & Stoller (2002), readers should have necessary ability to comprehend a text, to process the information, and to interpret appropriately what the writer intends. However, Nunan (2003) stated that reading comprehension is a fluent process by which information from the text with existing information can be combined to get a meaning. Moreover, comprehension is not only a linguistic skill, but also a general cognitive skill (Walter, 2007).

A variety of different factors can affect test takers' performance on any language test (Bachman, 1990). "If we are to develop language test appropriately, for the purposes for which they are intended, we must base them on clear definition of both abilities we wish to measure and the means by which we observe and measure these abilities" (p.81). Test method facet can be regarded as one of the most important factors which affect testee's performance. According to Bachman (1990), test method can be defined as "the characteristics of methods used to elicit test performance" (p.111). Bachman (1990) categorized test method facets into five...
major groups. They included (1) testing environment; (2) test rubrics; (3) the nature of the input; (4) the nature of the expected response; (5) the interaction between the input and response.

Bachman (1990) concluded that testee's performance is influenced by the characteristics of test method, and those constructed response types are commonly more difficult than selected ones. Accordingly, Rahimi (2007) pointed out that "a test used to assess a particular ability would yield different results when different test methods are used to gauge the same trait". As it seems obvious, the way reading comprehension is measured can be considered as another controversial issue. Because of a dynamic and multifaceted nature of it, reading comprehension seems too difficult to measure.

Allington (2001) held that reading comprehension has often been assessed by the ability of language learners who are able to recollect the details of what they have read. McKee (2012) maintained that a student can be accounted as a proficient reader by means of this kind of testing since they are able to answer the factual questions.

Reading comprehension including multiple components should be measured through different ways. Nevertheless, measurement of this type of skill depends on the mode of assessment and the format of the reading materials (Pearson & Hamm, 2005; Storch & Whitehurst, 2002).

One single type of measurement on reading comprehension ability cannot justify and assess the test taker's performance. Therefore, multiple measures may be necessary to gauge this type of skill (Falke, 2008). Besides, Flethcher (2006) claimed that a one-dimensional assessment for reading comprehension is innately imperfect. There is no one special type of a test that can consider all assessment dimensions. Multiple choice tests can be easily scored regarding accuracy and objectivity. In contrast to essay question tests, multiple choice tests cannot distinguish poor readers from competent and sophisticated ones appropriately (Chan & Kennedy, 2002). Furthermore, Chen (2010) asserted that subjectivity of the testers can produce an important effect on testee's scores leading to the problems of content validity. Dennis (2009) suggested that testers can rather make a fairer assessment of reading comprehension by employing a variety of diverse forms of tests.

A large number of researchers and authors recognized multiple choice questions as the most common measure of reading comprehension in research and utilized them in the standardized tests (Jenkins & Fuchs, 2003; Lahey, McNess, & Brown, 1973; Nesi et al., 2006; O'Connor & Klein, 2004; Frith & Snowling, 1983; Nourby & Bishop, 2002). However, reading comprehension can be measured through other types of test methods including cloze reading activities, gap filling, maze passages, and true-false.

Kobayashi (2002) investigated the effect of test formats on the students test performance. The test formats included cloze, open-ended question, and summary writing. The author found that test formats involved recall, cloze, maze, and multiple choice questions. This research was done on 164 fifth graders. Each student should read four texts. The results indicated that the readers performed significantly worse on recall test than the other types of reading comprehension tests. Moreover, the students performed best on multiple choice and maze tests. The author concluded that no one-sided and one-dimensional measurement can assess reading comprehension properly. But all four types of measurement constructed a more comprehensive and instructive picture.

Kinstsch and Yarbrough’s (1982) studied the effects of two types test formats on reading comprehension: cloze test and open-ended questions. The results showed that open-ended questions measured the main ideas comprehension of the text more effectively than cloze tests. In other words, cloze tests did not manifest the reader's overall understanding.

Shohamy (1984) examined whether testing formats - multiple choice and open ended questions - could make a difference on the readers' performance of reading comprehension or not. She found that significant differences between the performances of low-proficient and high-proficient readers. She came to this conclusion that multiple choice questions were easier than open ended questions. Chen & Cao (1999), however, stated that short answer questions were more effective in comparison to multiple-choice questions in testing of reading comprehension.

Liu (1998) carried out a study on test methods of reading comprehension. He employed three test formats including multiple choice questions, true-false questions, and short answer questions. The results revealed that the subjects' performance on reading comprehension tests was influenced by test methods. More specifically, test methods affect the high-proficient students more easily than low-proficient ones.

Sun (2001) used the above mentioned three test methods of reading comprehension test. Nevertheless, she reached a different conclusion. Indeed, results showed that there were no significant differences among the three mentioned test methods.

Liu (2009) conducted an experimental research to discover the effect of three different test methods on reading comprehension. The author found that three test methods including multiple-choice questions, gap filling, and short answer questions have a significant effect on reading comprehension. Results indicated that multiple choice questions and short answer questions were easier whereas gap filling was the most difficult.

The primary focus of earlier studies and experiments has been on the performance of the students, particularly the university students. Based on the earlier studies, the current study aims to assess and estimate
the effect of item variety on the performance of students’ reading comprehension. However, the researcher has found few studies conducted on high school students or EFL learners in institutes of Iran. For this reason, the researcher made attempt to investigate the impact of three test varieties on the performance of Iranian EFL learners’ reading comprehension in an English institute of Iran. To this aim, the present study attempts to address the following question: What is the effect of item variety on the performance of EFL learners’ reading comprehension?

METHODOLOGY

Participants
The participants in the present study were 30 EFL learners (at elementary level) from one language institute in Kashan, Iran (Ofogh Language Institute). They were from two available classes the researcher had access to. Moreover, all subjects were female. Almost all of them were between 15 and 19 years of age. More specifically, they were high school students. The book taught for elementary level was “Four corners 2” written by Jack. C. Richard & David Bohlke.

Instrument
Three reading passages were used in the current study; all of them contained the same general topic of communication. The passages, in more detail, were under the titles of “Baby talk”, “Face reading”, and “A language for women only”, respectively. The passages were chosen from a book entitled "Read This 1" by Daphne Mackey. The purpose for choosing the mentioned book lies in the fact that "Read This 1" was designed for elementary level English learners. To ensure suitability of three mentioned above reading passages to the reading texts of “Four Corners 2”, the readability of reading passages was measured.

Each of the three reading passages was followed by three types of questions including 5 true-false questions, 7 multiple-choice questions, and 7 gap filling questions, respectively. Totally, the participants were asked to answer 19 questions for each reading passage. Multiple-choice questions contained three options. Furthermore, seven words were presented for gap filling section. That is, no extra words were given for being chosen by the participants. For each blank, there was just one correct word.

Moreover, the statistical SPSS 21.0 for windows was used to analyze the data in this study.

Procedures
Three passages were distributed among the participants. More specifically, one reading passage was given to the subjects in one session. The subjects were asked to read the passages attentively and then respond to the three types of questions. The time limit allocated for reading passages and answering the questions was about 30 minutes.

In the next step, the reading passages were collected in order to be corrected. One mark was given for every correct answer, and zero if wrong. The total score was 19 for every passage and 57 for three passages.

Data analysis
Test scores on the reading comprehension from three item varieties (three formats) - true/false, multiple choice, and gap filling- were obtained. More specifically, the scores of the students in three test formats were summed. Then the total scores of each item variety were calculated. It is worth noting that the scores of each three test formats were calculated based on percentage, because the number of questions in true/false format was different from other two test formats (7 item for multiple choice or gap filling). For example, if one participant received the perfect score (answered all questions correctly) the calculated score for her was 100.

Since three test methods (item variety) were adopted on just one group of EFL learners, the researcher employed a dependent test (dependent significant test) for data analysis. Moreover, “three item varieties” (more than two types of mean scores) were used in order to compare and measure the students’ performance on reading comprehension. For this reason, repeated measures ANOVA (Analysis of Variance) were applied.

RESULT AND DISCUSSION

As mentioned earlier, the purpose of the present study was assessing the effect of item varieties on the performance of EFL learners’ reading comprehension.

Table 1 displayed both mean scores and standard deviations of the three test methods (true/false, multiple choice, and gap filling) in reading comprehension.

<table>
<thead>
<tr>
<th>Test method</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/False</td>
<td>74</td>
<td>23.64</td>
</tr>
<tr>
<td>Multiple choices</td>
<td>68.25</td>
<td>18.30</td>
</tr>
</tbody>
</table>
As pointed out in Table 1, multiple choice questions have the lowest mean score (68.25), whereas the participants obtained their highest mean scores in the gap filling questions (80). Furthermore, gap-filling questions have the highest standard deviation (26.74), which indicated that the scores were more widely spread. Overall, mean scores from the highest to the lowest belonged to gap-filling, true/false, and multiple-choice questions, respectively. Nevertheless, Table 1 did not indicate the fact whether the difference among mean scores were significant or not. Indeed, it did not show which method differs from which.

In order to assess the significance of difference among mean scores, repeated measures F test was used. Table 2 revealed the summary of repeated measures F test.

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test types</td>
<td>2069.86</td>
<td>2</td>
<td>1034.93</td>
<td>9.33</td>
<td>0.001</td>
</tr>
<tr>
<td>Error</td>
<td>6431.42</td>
<td>58</td>
<td>110.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As it was shown in Table 2, there was a significant difference among the mean scores elicited by the three test methods. However, as indicated by Table 2, it was not obvious which two test methods had a significant difference. In other words, the question of "Is there any significant difference between each pair of test methods? (TF & MC; TF & GF; MC & GF)" could not be answered according to Table 2. Therefore, the researcher employed LSD (Least Significant Difference) post hoc test to perceive any possible significant difference between each pair of two test methods.

Table 3 presented summary of Least Significant Difference post hoc test. Indeed, mean differences, standard error, and p-value were displayed in Table 3.

<table>
<thead>
<tr>
<th>Test type I &amp; J</th>
<th>Mean differences</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF &amp; MC</td>
<td>5.74</td>
<td>2.49</td>
<td>0.029</td>
</tr>
<tr>
<td>TF &amp; GF</td>
<td>-6</td>
<td>2.55</td>
<td>0.026</td>
</tr>
<tr>
<td>MC &amp; GF</td>
<td>-11.74</td>
<td>3.06</td>
<td>0.001</td>
</tr>
</tbody>
</table>

According to Table 3, mean differences between true/false & multiple choice questions were 5.74. Moreover, mean differences between true/false & gap-filling questions were -6; while this quantity was -11.74 between multiple choice & gap-filling questions. More specifically, the mean differences of true/false item variety, in comparison to multiple choice questions were significant (p<0.05). Besides, the mean differences of true/false item variety, in comparison with gap-filling questions were significant (p<0.05). Finally, the mean differences of multiple choice item variety compared to gap-filling were significant (p<0.01).

In sum, the results showed that three different item varieties (true/false, multiple choice, and gap-filling) made significantly different performances on reading comprehension of EFL learners. Indeed, differences in performance were significant on the three test methods. Findings indicated that the highest difference of performance (mean difference) can be pertinent to gap-filling, and the least difference of performance to multiple-choice. In other words, the most difficult item variety has been multiple-choice, whereas the easiest test method was gap-filling.

Liu (2009) argued that candidates had capability to employ some test taking strategies including guessing and deduction. He furthered what the candidates might need to do was to select one option in multiple-choice questions. That is, the students concentrated more on the provided options. Therefore, shorter time was in need to complete multiple-choice questions in comparison to other types of questions such as short answer items. The results, in his research, demonstrated that multiple-choice questions were easier. Moreover, many scholars maintain that multiple-choice questions evaluate or gauge only recognition knowledge. Candidates may choose an option correctly even though they don’t understand it properly. Also, Zheng, Cheng, & Klinger (2007) conducted an investigation into whether test formats in reading comprehension affected second-language students’ test performance differently. The results revealed that multiple-choice questions were generally considered to be easier to answer correctly than other types of test formats.

The findings in this study, however, were not in consistent with the results obtained by the two preceding mentioned experiments. The students in the current study found the multiple-choice questions most difficult. A likely explanation for this would be that nowadays, L2 reading comprehension test methods have been increasingly shifted from recognition or multiple-choice tests towards problem-solving tasks. Indeed, test constructors have recently attempted to concentrate on test items which may actively involve the students in analyzing, processing, and responding to the reading comprehension questions. As discussed earlier, multiple-
choice questions had been recognized as one of the most common and appropriate means for assessing the student’s knowledge of L2 reading comprehension for a long time. However, by giving the less attention to multiple-choice items, today’s students seemingly intend to engage with the kinds of the reading comprehension items or tasks which can further reinforce their cognitive abilities and strategies. That is, one possible justification for the discrepancy between this study and the prior research may be pertinent to focusing the recent test methods on processing and analyzing items rather than choosing the options by chance. Accordingly, teaching approaches and methods have strongly emphasized the importance of communicative goals in a few recent decades. It is apparently through communicative activities and functions that learners broaden and deepen the capacity for reading comprehension. Consequently, EFL learners appear to perform more efficiently and properly on the other types of items such as gap filling and open ended tests.

The next inference drawn from the results may lie in the fact that the participants in this study achieved the poorer results in multiple choice items due to the effect of chance. In other words, it seems that students choose the wrong options on multiple-choice questions because they may guess one option incorrectly and by chance.

On the other hand, some researchers in the field of test variety in reading comprehension have claimed that gap-filling tests seem to assess reading comprehension at a sentence level. In fact, gap-filling tests may be essentially incorporated in sentence bound category. For instance, Liu (2009) stated that “items in gap-filling tests can be answered without referring to the rest part of the passage” (p.150). One possible justification for the results of gap-filling test in the current study would be that students may choose the words from the provided list regardless to the passage. The participants, in this study, made possibly attempt to fill in the blanks regarding the words provided by the text constructor. They might ignore the content of the passage and rely only on the sentence. For this reason, they could choose the words correctly in this type of questions. Some students seem to be confused when they refer to the passage in order to complete the gaps. When they are bound to an only one sentence, they can probably select the options more freely and even correctly. As a result, the candidates in this study seemingly performed better on gap-filling tests rather than other kinds of items.

Besides, the findings showed that students in the current study found true-false questions easier than multiple-choice items and more difficult than gap-filling questions. Here, again, the issue of chance may play an important role on performance of students on choosing a dichotomy. One possible difference between the performance on true/false and multiple-choice items could be that the test takers had a probability of 50% for choosing correctly for true/false items; while the probability of selecting for multiple-choice items was around 33%.

CONCLUSION

The present research was intended to investigate how Iranian EFL learners perform on the reading comprehension with respect to item varieties. More specifically, the researcher made attempt to find the possible effects of three item varieties on performance of EFL learners in L2 reading comprehension. All in all, the results of the study mainly indicated that each item variety had a different and significant impact on the performance of participants’ reading comprehension. Indeed, differences in performance were significant on the three item varieties and students were affected by the item varieties. Furthermore, it was found that the subjects detected the multiple-choice items as the most difficult and the gap-filling questions as the most easiest.

According to Alderson (2000), it seems insufficient to test the performance and understanding of reading comprehension ability through just one method. Good reading comprehension tests involve a number of various items and methods. Moreover, Rubin (2011) suggested that different types of tests should be employed for assessment purposes by teachers. He furthered that teachers should be encouraged to utilize various test methods in order to make a more comprehensive evaluation of readers' comprehension. Dennis (2009) also emphasized that teachers using this approach would be able to detect the specific strengths and weaknesses in students’ reading comprehension more effectively. Consequently, it would be expected that we as teachers and simultaneously as test constructors tried our best to give the priority to use of different test formats or item varieties in reading comprehension measurement. By this, more effective and better results would be obtained in the field of assessing L2 reading comprehension.

REFERENCES


Rubin J.2011. Organizing and evaluating results from multiple reading assessments. The Reading Teacher, 64(8), 606-611.


