Assessing Reading Comprehension of Expository Text across Different Response Formats

Masoumeh Akhondi
Islamic Azad University, Malayer Branch, Iran
akhondi@iau-malayer.ac.ir

Faramarz Aziz Malayeri
Islamic Azad University, Malayer Branch, Iran
azizmalayeri@iau-malayer.ac.ir

Abstract

This study investigated if different response formats (test methods) measure reading comprehension of expository text differently. The study was conducted with 48 semester 6 TESL students at a university in Selangor, Malaysia. These students received an expository passage having descriptive rhetorical structure followed by three response formats, namely, incomplete outline, graphic organizer, and summary writing. Results from Repeated Measures two-way ANOVA indicated that high-achievers outperformed intermediate- and low-achievers across the three response formats. Moreover, the three groups achieved higher scores on incomplete outline and summary writing, respectively. However, graphic organizer appeared as the most difficult task as the respondents achieved the lowest score in this task.

Keywords: Assessing Reading, Response Formats, Expository Text, Reading Comprehension

Received: February 2010; Accepted: December 2010

1. Introduction

Reading is not a unifaceted skill that can be measured by a single test method. It is a multi-component ability which calls for a variety of response formats (test methods) to measure this multivariate skill properly (Akhondi, 2011).

Reading as a multi-factor skill involves a complex combination and integration of a variety of cognitive, linguistic, and non-linguistic skills (Nassaji, 2003). Among the many existing variables that are considered affecting language test performance, one central issue is the effect of response formats (test methods) on test performance (Alderson, 2000; Bachman and Palmer, 1996; Brantmeier, 2006; Buck, 2001). There has been an ongoing debate in the reading research literature for the last two decades to locate the best response format (test method) as the best instrument to measure the knowledge of text structure or reading comprehension in all.

Despite the fact that the importance of the knowledge of text structure is widely recognized in reading research and instructional practice, it is not yet clear in reading assessment research how best to assess this knowledge (Alderson, 2000). Fletcher (2006) identified three issues that are practically important for the measurement of reading comprehension, namely: the nature of the text, how reading comprehension is assessed, and individual differences. Many factors affect success or failure during comprehension of expository texts. Reader's characteristics, background knowledge, text properties, awareness of text structure, and the instructional context in which reading takes place are just a few (Carrell, 1992; van den Broek, Fletcher, and Risden, 1993; van den Broek and Kremer, 1999; Williams, 2005; Nassaji, 2007; Kendeou and van den Broek, 2007). Although these factors have often been studied in isolation, it is their interactions and interdependencies that provide important information about naturalistic text comprehension (Raap and van den Broek, 2005).

In a study, Kendeou and van den Broek (2007) investigated the effect of prior knowledge and text structure on cognitive processes during comprehension of scientific texts. They found out that readers adjust their processing as a function of the interaction between prior knowledge and text structure. Furthermore, the results showed that readers' memory for the text was affected by differences in their prior knowledge, independently of text structure. As Pearson and Hamm (2005) summarized, early research identified that reading comprehension involved multiple components that would appear depending on the formats used to present the material to be read and the manner in which the person was asked to indicate their understanding of the material that was read. As Francis et al. (2006) argue any single one-dimensional attempt to assess reading comprehension is inherently imperfect.

From the psychometric view of Francis et al., differences across methods used to measure reading comprehension can be interpreted as degrees of how well different indicators identify one or more latent variables that make up reading comprehension. Johnson, Jenkins, and Jewell (2005) have emphasized the same viewpoint. They believe that reading is a multicomponential ability, which needs to be assessed by a multidimentional test or a set of test formats each of which assesses one dimension of reading ability. Williams et al. (2009) argue that the text structure program is successful when comprehension is evaluated on tasks requiring both spoken and written responses.

Despite this historical emphasis, many modern approaches to the assessment of reading are one-dimensional, with little variation in the material the person reads and relatively narrow response formats that do not vary within the test (Fletcher, 2006). Different test formats such as summary tasks, outlines, and graphic organizers have been proposed for use in reading classrooms to enhance learners' awareness of text structures. These test

formats assess the extent to which students have achieved knowledge of text structure (Grabe and Stoller, 2002). To assess reading comprehension, some tests rely almost exclusively on multiple-choice, others on fill-in-the-blank (cloze), and others on retells. Many reading research studies investigating the nature of this knowledge have used recall protocols as a measurement instrument (Aulls, 1975; Carrell, 1992; Meyer, Brandt, and Bluth, 1980). The drive for high reliability, especially on high-stakes assessments, often leads to significant restrictions of both the type of material that must be read and the response formats. As Cutting and Scarborough (2006) believe, the inferences that are made about how well an individual person comprehends written material vary depending on how it is assessed. As Fletcher (2006) believes, the assessment of reading comprehension is difficult because it is not an overt process that can be directly observed. Rather, only the products of the process of comprehension are observed, and an inference is made about the nature of the processes and the quality of comprehension. Assessing reading comprehension is challenging, because it is a complex and multiply determined outcome (RAND Reading Study Group, 2002). The RAND Group also believe that inferring how well a person comprehends is the real problem in measuring reading (and language) comprehension. This would be achieved when the researcher uses different kinds of texts and test formats together, because each of these measure different aspects of reading ability and sources of reader's failure in reading will be distinguished (Fletcher, 2006).

With the acknowledgement to these efforts, it is necessary to note that there have been few attempts to compare the effectiveness of the three response formats (incomplete outline, graphic organizer, and summary writing) in measuring the knowledge of text structure. It, therefore, remains to be investigated whether these three response formats reveal different results of

the comprehension of the same expository text at different levels of proficiency. Hence, the present study addresses the following research question: do the students of similar proficiency perform differently in different response formats?

2. Expository Text Structure

Expository or "informational" texts convey and communicate factual information. These texts contain more unfamiliar vocabulary and concepts, and fewer ideas related to here-and-now, and less information directly related to personal experience (Hall et al., 2005). Both narrative and expository texts have hierarchical structures (Meyer and Poon, 2001); however, narrative texts tend to follow one structural pattern (story grammar), whereas; multiple text structures are used in expository texts. Basic structural patterns include description, sequence, problem/solution, compare/contrast, and cause/effect (Meyer and Freedle, 1984). The bulk of previous L2 research on the knowledge of expository text structure revealed that among the five text structures proposed by Meyer and Feedle (1984), cause/effect appears as the most difficult rhetorical structure to be identified by the respondents, while description and sequence structure appear as the easiest structure to be probed. Meyer's method of text structure analysis, or the content structure analysis model (Meyer, 1975), arranges the information in a text into a hierarchy. The ideas at the top of the content structure tree represent the main ideas of the text while the ideas at the bottom of the tree represent details. Competent readers are normally able to identify the main idea of the passage followed by the major ideas, and the related supporting details. The proficient readers are successful in discovering the underlying rhetorical structure of the passage, different levels of ideas, and their interrelationships properly.

3. Response Formats

Response format is defined as: "the way in which the response is produced, for instance selected, limited production, and extended production (short answer, summary writing, multiple choice, etc.)" (Bachman and Palmer, 1996). These different response formats have all been used in measuring reading comprehension as a major module of language ability. Bachman's Models of Language Ability (Bachman, 1990; Bachman and Palmer, 1996) have emphasized the role of response format (expected response). He maintains that these characteristics constitute a major theoretical foundation for his model.

Reading researchers have argued that different test formats seem to measure different aspects of language ability (Graesser, Hoffman, and Clark, 1980; Kintsch and Yarbrough, 1982; Lewkowicz, 1983; Reder and Anderson, 1980; Shohamy, 1984; Shohamy and Inbar, 1991; Kobayashi, 1995, 2002; Cutting and Scarborough, 2006; Francis et al., 2006). The multiple-choice format, for example, has been heavily criticized because it seems that test takers can guess the right answer without fully understanding the reading passage (Nevo, 1989; Katz, Lautenschalger, Blackburn, and Harris, 1990; Royer, 1990; Weir, 1993). The use of alternatives to multiple-choice formats has been proposed by language testers (e.g., Brown and Hudson, 1998) and their test validity has been examined. Kobayashi (1995, 2002) scrutinized the effects of text organization and response format on second language learners' performance in reading comprehension tests. She argues that there is interaction between text organization and test format and this interaction significantly affects students' performance on reading comprehension tests. Moreover, she continues to argue that more proficient learners achieved higher scores in summary writing and open-ended questions when texts were more

clearly organized. Many studies on reading assessment experienced different techniques –with their potential merits and demerits– which might be used to assess reading comprehension; however, Alderson (2000) believes that: "no single test method can fulfil all the varied purposes for which we might test" (p. 203). He continues to argue that it would be naive to assume that because a test method is widely used it is therefore 'valid'. This conclusion can be drawn out of Alderson's discussion on test method that it is wise to examine all the research in that specific field and not that which shows the benefits of a given method. Moreover, it would be more reasonable to use multiple measures (test methods) to assess an ability in order to draw more valid and accurate results of the variable of interest because each test method measures a specific aspect of language ability.

4. The Purpose of the Study

The present study is planned based on an *ex post facto* design in the way that the researchers manipulated the response formats and investigated the probable consequences of this manipulation on students' reading comprehension performance. The participants have already been measured on their reading proficiency through a standardized reading comprehension test in order to make their performance comparable in three levels of proficiency across the three response formats. In other words, the students' ranking from the high-achievers to low-achievers is applied to discuss the mean differences of three response formats between the three groups of skilled to less-skilled students. The current study addresses the following research question: *do the students of similar proficiency perform differently in different response formats?* This main research question could be expanded into three specific research questions as follows:

1. Is there any significant main effect of reading proficiency on the students' performance across the three response formats? (Treatment main effect)
2. Is there any significant main effect of response formats (Test method effect) on the students' performance across the three response formats? (Trial main effect)
3. Is there any significant interaction effect of response formats and reading proficiency on the students' performance across the three response formats? (Interaction effect)

5. Methodology

5.1. Participants

The participants in this study were 48 undergraduate TESL students. The researchers selected these students because it is hypothesized that they have had the maximum exposure to the academic reading texts and their reading ability had already been developed to some extent. As Hall et al. (2005) argue, the amount of experience that the students receive with expository texts certainly impacts their ability to deal effectively with this type of text. Substantial experience with a genre is necessary for knowledge of that genre to develop (Duke, 2000).

5.2. Procedures

The participants completed the reading tasks in a single session that lasted for approximately 2 hours. The reading passage was about the initial period 1800-900 B.C. This passage was a descriptive passage in which it discussed the ways the Indian American changed their way of life in this period. The tasks were ordered the same way for all of the participants, having the summary writing as

the first task followed by graphic organizer and incomplete outline. The reading proficiency test had been used to assess the students on their academic reading ability in individual session prior to the session they took the three test tasks to measure the text structure knowledge and reading comprehension of the expository text.

5.3. Materials

5.3.1. Reading Proficiency Test

In order to make the participants' performance comparable across the three response formats, a standard reading proficiency test (MUET), Malaysian University English Test, comprised of 50 multiple-choice reading comprehension items was administered. MUET has been standardized within Malaysia and its validity has been established since its debut in 2000. Ever since, the test bank has been updated, calibrated and refined to assure dependable test scores for the Malaysian tertiary education (Arshad, 2004). Statistical analysis confirmed that there was a significant difference among the respondents in three levels of reading proficiency [F (2, 45) = 3.875, P= .028]. Furthermore, the results of post-hoc comparison revealed that the significant difference could be found between the high-achievers and low-achievers (P=.008), but no statistical significant difference existed between the high- and intermediate-achievers (P=.05), as well as the intermediate- and low-achievers (P=.212).

5.3.2. The Summary Task

The first task was a summary task. In this task, examinees wrote a summary of the text in English. The examinees were asked to write a well-organized paragraph that

was no longer than 200 words. They were also told that their summaries should not exceed the space provided on the answer sheet. The instructions of the task stated that the summary should include the overall main idea of the passage, the major ideas that the passage covers, and the supporting details of the major ideas. The examinees were also told in the task instructions that their summary would be scored based on how well they have recognized the hierarchy of the ideas in the text, not on grammatical accuracy.

5.3.3. The Graphic Organizer Task

The second task was a graphic organizer task. In this task, examinees filled in a table with their own words about the overall main ideas of the reading passage, the major ideas that the passage covered, and the supporting details that support those major ideas. The examinees put each major idea in the left column and then put the supporting details of that major idea in the right column. The answers could be written in phrases or in complete sentences. It was hypothesized that the examinees need to be sensitive to the hierarchy of ideas presented in the reading passage and be able to distinguish different levels of ideas (the overall main idea, the major ideas, and the supporting details) in order to successfully perform this task.

5.3.4. The Incomplete Outline Task

The final task was an incomplete outline task. In this task, the examinees completed a partially blank outline with appropriate major ideas and supporting details from the reading passage. The hierarchy of the ideas in the outline was signalled by Roman numerals, capital letters, and Arabic numerals. Each notation represented levels of ideas ranging from the most to the least important. In other words, Roman numerals

represented the most important main ideas while the capital letters represented the major ideas and Arabic numerals represented the supporting details in the text. It was hypothesized that the examinees need to use their knowledge of text structure to determine the answer for each blank on the outline.

6. Data Analysis

In order to test the hypotheses formulated on the differences across the three levels of reading proficiency and the three response formats, the test of General Linear Model (GLM) Repeated Measures Two-Way ANOVA was used. The hypothesis formulated for the present research tested the significant difference between the three groups of achievers in their performance in different reading comprehension test formats. It is assumed that the test is capable enough to reveal the variation observed across the 3×3 interrelationships across the levels of the Independent and the Dependent variables. The rationale behind using Repeated Measures two-way ANOVA is to precisely locate the differences of participants' performances due to their levels of proficiency across the three test formats. However, the descriptive statistics for the 3×3 levels of the factors reveal the differences among the three proficiency levels across the response formats.

Table 1: Descriptive Statistics (Proficiency levels* Test formats)

Test Formats	Proficiency Level	Mean	Std. Deviation	N
Incomplete Outline	1	90.25	11.67	8
	2	73.26	19.45	26
	3	70.64	21.28	14
	Total	75.33	19.84	48
Graphic Organizer	1	67.12	19.23	8
	2	64.65	19.63	26
	3	58.92	11.57	14
	Total	63.39	17.50	48
Summary Writing	1	80.37	12.21	8
	2	68.65	16.14	26
	3	60.92	15.43	14
	Total	68.35	16.36	48

1= high-achievers 2= intermediate-achievers 3=low-achievers

Results from descriptive statistics indicated that the mean score for high-achievers on incomplete outline test (N=8, $\underline{M}$=90.25, SD=11.67) was significantly different from the two other groups of intermediate-achievers (N=26, $\underline{M}$=73.26, SD=19.45) and low-achievers (N=14, $\underline{M}$=70.64, SD=21.28). Comparing the results of incomplete outline task with the other two tasks revealed that the three proficiency groups performed well on incomplete outline task, while the weakest performance for the three groups was observed on graphic organizer task ($\underline{M}_{high}$=67.12, SD_{high}=19.23; $\underline{M}_{intermediate}$=64.65, $SD_{intermediate}$=19.63; $\underline{M}_{low}$=58.29, SD_{low}=11.57). If we consider these three response formats on a continuum, incomplete outline task and graphic organizer task could be placed at either extreme of this continuum. In other words, at one extreme, there is incomplete outline task with the highest performance mean score for the three groups while at the other extreme of this continuum graphic organizer task is located with the lowest

performance mean score for the three groups. However, the summary writing task stands somewhere in the middle of this continuum with medium mean score for the three groups ($\underline{M}_{high}$=80.37, SD_{high}=12.21; $\underline{M}_{intermediate}$=68.65, $SD_{intermediate}$=16.14; $\underline{M}_{low}$=60.92, SD_{low}=15.43).

Mauchly's test of sphericity is used to test meeting the assumptions of univariate models and tests of within-subject effects, as illustrated in Table 2.

A finding of non-significance value (P=.121>.05) corresponds to concluding that sphericity assumption is met; therefore, we can use the pooled test which is more powerful compared to the other tests.

Table 2: Muchly's test of sphericity

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
Test formats	.908	4.225	2	.121

Due to the fact that the sphericity assumption is met, the values for the pooled test (Sphericity Assumed) are reported in Table 3 below.

Table 3: Tests of within-subjects effects (Pooled test of test format effect)

Sources	Type III Sum of squares	df	Mean square	F	Sig.	Partial η^2	Observed Power
#Test formats	4036.976	2	2018.488	9.607	.000	.176	.978
#Test formats* Proficiency level	767.479	4	191.87	.913	.460	.039	.279
#Error (test formats)	18910.42	90	210.116				

α=.05 η2= .o1 (small effect size) η2= .o6 (moderate effect size) η2= .14 (large effect size), # Sphericity Assumed

As the values from the pooled test Table indicate, there are significant differences [F (2, 90) =9.6, P=.000] in three test formats. This finding supports the descriptive statistics Table which was reported earlier in this paper. It could be interpreted that the participants in three levels of proficiency performed well in the incomplete outline task and the weakest performance was observed

in the graphic organizer task (See Table1 for further information). The effect size (partial eta squared) for the significant main effect of test formats on students' performance is considered large (η^2=.176). The results for the interaction effect of test formats and proficiency level indicated no significant difference [F(4, 90)=.913, P=.460]. Moreover, the effect size (partial eta squared) for the interaction effect of test formats and reading proficiency is considered small (η^2=.039). This could be discussed in this way that the three proficiency levels performed similar to each other across the test formats; to put it another way, the participants in the three proficiency levels achieved the highest mean scores in the incomplete outline task, while, the lowest mean performance for the three groups was observed in the graphic organizer task and a medium performance mean for the summary writing. Moreover, the Box's test of equality of covariance matrices appeared to be non-significant (P=.221); therefore, the assumption of equality of variance and covariance across each cell in the Table was held. This finding is also supported by Levene's test of equality of variances. The results from the Levene's test indicated non-significant values for the three test formats ($P_{Incomp.}$=.209; $P_{Graphic.}$=.110; $P_{Summ.}$=.6); in other words, this proves the equality (homogeneity) of variances across test formats. This is good news for the researchers; hence, the results from Multivariate Test (Table 4) could be reported to support the findings from other statistical tests which were reported earlier in this paper.

Table 4: Multivariate tests[d]

Effect		Value	F	df	Sig.	Partial η^2	Observed Power
Test formats	Wilks' Lambda	.737	7.854	2	.001	.263	.939
Test formats* proflevel	Wilks' Lambda	.924	.891	4	.473	.039	.273

α=.05 η^2= .o1 (small effect size) η^2= .o6 (moderate effect size) η^2= .14 (large effect size)

As it is depicted in Table 4, the value of Wilks' Lambda for the test formats (Trial Main Effect) appeared as a significant value (F=7.854, P=.001). Here, the significance value of Wilks' Lambda indicated a significant Trial Main Effect in which it provides the answer to the second research question of the current research [*Is there any significant main effect of response formats (test method effect) on the students' performance across the three response formats? (Trial main effect)*]. Moreover, the effect size for the main effect of test formats is considered large (η^2=.263). It is interpreted that there is a significant effect of response formats on students' performance across these formats, or, to put it other way, the students did perform differently in each of the response formats. It could be interpreted that, the incomplete outline was the easiest task for them to complete while the graphic organizer was a difficult task to accomplish. In contrast, the value of Wilks' Lambda revealed a non-significant interaction effect of response formats and proficiency levels (F=.891, P=.473). Besides, the effect size for the interaction effect of test formats and reading proficiency is considered small (η^2=.263). Therefore, the third research question [*Is there any significant interaction effect of response formats and reading proficiency on the students' performance across the three response formats? (Interaction effect)*] is answered in a way that the respondents treated the response formats similarly regardless of their level of proficiency. It means that for the three proficiency levels, incomplete outline was the easiest task and the graphic organizer due to the task unfamiliarity was the most difficult task and the summary writing task was a task of medium difficulty to all proficiency levels.

Table 5: Tests of between-subjects effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Proficiency level	3793.44	2	1896.72	3.875	.028	.672
Error	22027.782	45	489.506			

α=.05

By scrutinizing the Between-Subject Effect Table, the answer to the first research question in the current research will be provided [*Is there any significant main effect of reading proficiency on the students' performance across the three response formats? (Treatment main effect)*]. Test of Between-Subject Effects revealed that there is a significant treatment effect [F (2,45)=3.875; P=.028] of reading proficiency on students' performance across the three response formats. This means that the highest performance mean score in all of the three response formats belonged to the high-achievers, whereas the lowest performance mean score had been observed in the low-achiever group and the intermediate-achievers' performance mean stood somewhere in the middle point of the two other groups. This illustrates that as the reading proficiency increases the performance on the three response formats increases, vice versa.

Table 6: Pairwise Comparisons (Proficiency levels)

(I) Prof.level	(J) Prof.level	Mean Difference (I-J)	Sig.	95% Confidence Interval for Difference	
				Lower Bound	Upper Bound
1	2	10.391	.05	-.011	20.793
	3	15.750*	.008	4.347	27.153
2	3	5.359	.212	-3.170	13.888

1= high-achievers 2= intermediate-achievers 3=low-achievers

*.The mean difference is significant at the .05 level

The result of Pairwise Comparisons (post-hoc) for the proficiency levels revealed that there is a significant difference only between the high-achievers

and the low-achievers (P=.008), while no statistically significant difference had been observed between the other groups of proficiency.

Table 7: Pairwise Comparisons (Test formats)

(I) testformat	(J) testformat	Mean Difference (I-J)	Sig.	95% Confidence Interval for Difference	
				Lower Bound	Upper Bound
1	2	14.485*	.000	7.135	21.835
	3	8.068*	.023	1.143	14.993
2	3	-6.417*	.026	-12.028	-.805

1= Incomplete Outline 2= Graphic Organizer 3=Summary Writing

*.The mean difference is significant at the .05 level

The result of Pairwise Comparisons (post-hoc) for the test formats indicated significant differences ($p<.05$) among the three response formats (See Table 6 for further information). It is interpreted that the mean score for the three response formats statistically differs from each other.

As it is depicted in Figure 1, the interaction of proficiency levels and response formats has a clear cut indicator among the performance of the three proficiency groups. In this figure, it is illustrated that with the increase of academic reading ability (reading proficiency the test takers' performance on the three response formats (test formats) has increased. In the case of graphic organizer task, the proficiency level did not really matter as the lowest mean of the performance for the three groups was observed on this task. From the respondents' feedback, it was revealed that they were not familiar with this kind of test task and they failed to accomplish the task successfully but they felt easy to complete the incomplete outline task as some of the ideas were already completed. Moreover, some of the participants who thought that they were not good at writing proficiency were reluctant to write a summary out of the text they had read.

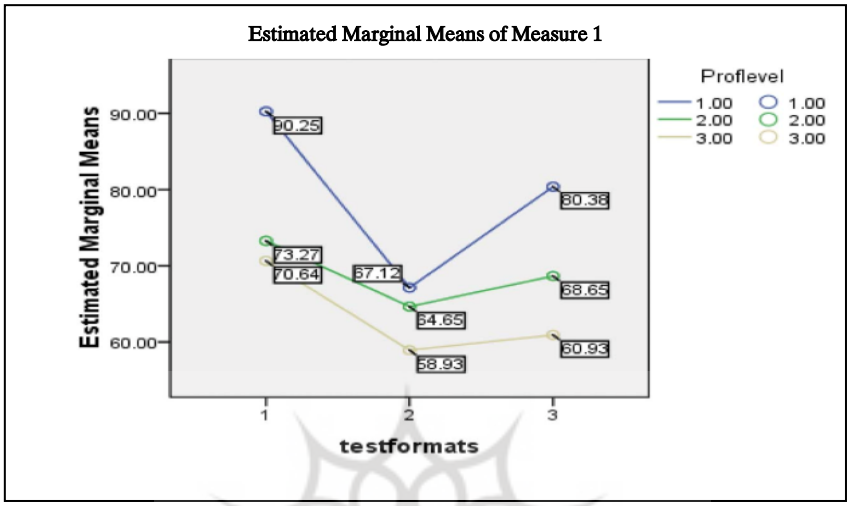

Figure 1. Interaction plot of proficiency levels and test formats

1. The means of the **incomplete outline task** for the three proficiency levels (high=90.25; intermediate=73.27; low=70.64)
2. The means of the **graphic organizer task** for the three proficiency levels (high=67.12; intermediate=64.65; low=58.93)
3. The means of the **summary writing task** for the three proficiency levels (high=80.38; intermediate=68.65; low=60.93)

*** Test formats (1= incomplete outline; 2= graphic organizer; 3= summary writing)*

7. Discussion

Reading researchers have argued that different test formats seem to measure different aspects of language ability (Graesser, Hoffman, and Clark, 1980; Kintsch and Yarbrough, 1982; Lewkowicz, 1983; Reder and Anderson, 1980; Shohamy, 1984; Shohamy and Inbar, 1991; Kobayashi, 1995, 2002; Cutting and Scarborough, 2006; Francis et al., 2006). The multiple-choice format, for

example, has been heavily criticized because it seems that test takers can guess the right answer without fully understanding the reading passage (Nevo, 1989; Katz, Lautenschalger, Blackburn, and Harris, 1990; Royer, 1990; Weir, 1993). The use of alternatives to multiple-choice formats has been proposed by language testers (e.g., Brown and Hudson, 1998) and their test validity has been examined. Many studies on reading assessment experienced different techniques –with their potential merits and demerits– which might be used to assess reading comprehension; however, Alderson (2000) believes that: "... no single test method can fulfil all the varied purposes for which we might test" (p. 203). He continues to argue that it would be naive to assume that because a test method is widely used it is therefore 'valid'. Moreover, it would be more reasonable to use multiple measures (test methods) to assess an ability in order to draw more valid and accurate results of the variable of interest because each test method measures a specific aspect of language ability.

Kobayashi (1995, 2002) scrutinized the effects of text organization and response format on second language learners' performance in reading comprehension tests. She argues that there is an interaction between text organization and test format and this interaction significantly affected students' performance on reading comprehension tests. Moreover, she continues to argue that more proficient learners achieved higher scores in summary writing and open-ended questions when texts were more clearly organized. Kobayashi's (2002) findings support the findings of the current research with respect to the respondents' proficiency level in summary writing task in which the skilled respondents outperformed the intermediate- and the low-achievers, respectively. However, Kobayashi's findings in the correlations coefficients between students' proficiency and the reading comprehension performance

were found contrary to the findings of the current research which indicated the null effect of proficiency on graphic organizer task.

In another study, Vongpumivitch (2004) conducted a research examining the nature of the knowledge of text structure by investigating the performance of ESL test takers on four test tasks designed to measure the knowledge of text structure: an incomplete outline, a graphic organizer, a summary writing, and a set of open-ended questions. She argues that each of the four tasks measured the knowledge of text structure to different extents. The finding of her study revealed that incomplete outline task was the easiest task of all, while the graphic organizer was located as the most difficult task. The findings of the present study are in line with those of Vongpumivitch's study (2004) in that the incomplete outline was always the easiest task, while the hardest task was the graphic organizer task. Carrell (1987b) agrees that outlines are popular in reading classrooms because of their emphasis on the hierarchy of ideas; a reader can easily return to the top-level ideas in a text by looking back at an outline. Contrary to the Carrell's positive view towards effectiveness of outline task, Tuckman (1993) in a study on first language reading found that outlining can be problematic because (a) many students were not used to create an outline of a reading passage, and (b) even after being taught about outlining, many students were still not able to make outlines appropriately.

Vongpumivitch (2004) also found a significant relationship between the knowledge of text structure and academic reading ability (reading proficiency), and language ability as a whole. This finding of Vongpumivitch's study supports the finding of the current research as the higher the reading proficiency levels, the better would be their performance on the test tasks. Kobayashi (2002) and Vongpumivitch (2004) agree on that, the summary writing task was the easiest

task [after the incomplete outline task] when the respondents were requested to locate the main idea of the reading task in hand.

Ozuru, et al. (2007) believe that the test format which is used to assess comprehension of expository text is of a paramount significance. They continue to argue that the same reader may appear to be a relatively strong or weak reader depending on a specific reading comprehension task used to assess his/her reading comprehension ability. Therefore, it is recommended to apply a variety of test formats to assess reading comprehension of expository text to come up with the most appropriate test method to be used in reading research and practice. Yet, there is a big gap of research interest in investigating the plausible relationship between test tasks and text structure across available research literature which calls for a closer scrutiny. The researchers would like to recommend the replication of the current research with different text structures and response formats in other ESL contexts. It is argued that variables like topical knowledge and the task familiarity would drastically affect the test takers' performance, hence, affect the validity and generalizability of the sampled performance.

References

Akhondi, M. (2011). Measuring the knowledge of expository text structure across different text types and response formats in an ESL academic context. Unpublished PhD dissertation, Universiti Putra Malaysia.

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

Aulls, M. W. (1975). "Expository paragraph properties that influence literal recall", *Journal of Reading Behavior*, 7(4), pp. 391-400.

Arshad abd. S. (2004). *Essentials of Language Testing for Malaysian Teachers,* Serdang: University Putra Malaysia Press.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*, Oxford University Press.

Bachman, L. F., and Palmer, A. S. (1996). *Language Testing in Practice,* Oxford: Oxford University Press.

Brantmeier, C. (2006). "Readers' gender and test method effect in second language reading", *Forum on Public Policy*, pp. 1-36.

Brown, J. D., and Hudson, T. (1998). "The alternatives in language assessment", *TESOL Quarterly*, 32(4), pp. 653-675.

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

Carrell, P. L. (1992). "Awareness of text structure: effects on recall", *Language Learning*, 42(1), pp. 1-20.

Cutting, L. E., and Scarborough, H. S. (2006). "Prediction of reading comprehension: relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured", *Scientific Studies of Reading*, 10(3), pp. 277-299.

Duke, N. K. (2000). "3.6 minutes per day: the scarcity of informational texts in the first grade", *Reading Research Quarterly*, 35, pp. 202-224.

Fletcher, J. M. (2006). "Measuring reading comprehension", *Scientific Studies of Reading*, 10(3), pp. 323-330.

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., and Iglesias, A. (2006). "Measures of reading comprehension: a latent variable analysis of the diagnostic assessment of reading comprehension", *Scientific Studies of Reading*, 10(3), pp. 301-322.

Graesser, A. C., Hoffman, N. L., and Clark, L. F. (1980). "Structural components of reading time", *Journal of Verbal Learning and Verbal Behavior*, 19, pp. 135-151.

Grabe, W., and Stoller, F. (2002). *Teaching and Researching Reading*, New York: Longman.

Hall, K. M., Markham, J. C., and Culatta, B. (2005). "The development of the early expository comprehension assessment (EECA): a look at reliability", *Communication Disorders Quarterly*, 26(4), pp. 195-206.

Hall, K. M., Sabey, B. L., and McClellan, M. (2005). "Expository text comprehension: helping primary-grade teachers use expository texts to full advantage", *Reading Psychology*, 26, pp. 211-234.

Johnson, E. S., Jenkins, J. R., and Jewell, M. (2005). "Analyzing components of reading on performance assessments: an expanded simple view", *Reading Psychology*, 26, pp. 267-283.

Katz, S., Lautenschalger, G., Blackburn, A., and Harris, F. (1990). "Answering reading comprehension items without passages on the SAT", *Psychological Science*, 1, pp. 122-127.

Kendeou, P., and van Den Broek, P. (2007). "The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts", *Memory and Cognition*, 35(7), pp. 1567-1577.

Kintsch, W., and Yarbrough, J. C. (1982). "Role of rhetorical structure in text comprehension", *Journal of Educational Psychology*, 74(6), pp. 828-834.

Kobayashi, M. (1995). Effects of text organization and test format on reading comprehension test performance. Unpublished PhD dissertation, Thames Valley University.

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

Lewkowicz, J. A. (1983). Method effect on testing reading comprehension: Acomparison of three methods. Unpublished MA Thesis, University of Lancaster.

Meyer, B. J. F. (1975). *The Organization of Prose and Its Effects on Memory*, Amsterdam: North-Holland Publishing Company.

Meyer, B. J. F., Brandt, D. M., and Bluth, G. J. (1980). "Use of top-level structure in text: Key for reading comprehension of ninth-grade students", *Reading Research Quarterly*, 16(1), pp. 72-103.

Meyer, B. J. F., and Freedle, R. O. (1984). "Effects of discourse type on recall", *American Educational Research Journal*, 21(1), pp. 121-143.

Meyer, B. J. F., and Poon, L. W. (2001). "Effects of structure strategy training and signaling on recall of text", *Journal of Educational Psychology*, 93(1), pp. 141-159.

Nevo, N. (1989). "Test-taking strategies on a multiple-choice test of reading comprehension", *Language Testing*, 6(2), pp. 199-215.

Nassaji, H. (2003). "Higher-level and lower-level text processing skills in advanced ESL reading comprehension", *The Modern Language Journal*, 87(2), pp. 261-276.

Nassaji, H. (2007). "Schema theory and knowledge-based processes in second language reading comprehension: a need for alternative perspectives", *Language Learning*, 57(1), pp. 79-113.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., and McNamara, D. S. (2007). "Influence of question format and text availability on the assessment of expository text comprehension", *Cognition and Instruction*, 25(4), pp. 399-438.

Pearson, P. D., and Hamm, D. N. (2005). "The assessment of reading comprehension: A review of practices – past, present, and future", (pp. 13-69), in Paris, S. G., and Stahl, S. A. (eds.), *Children's Reading Comprehension and Assessment.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers, pp. 131-160.

RAND Reading Study Group. (2002). Reading for understanding: Toward an R&D program in reading comprehension, Washington, DC: RAND Education.

Raap, D., and van den Broek, P. (2005). "Dynamic text comprehension: an interactive view of reading", *Current Directions in Psychological Science*, 14, pp. 276-279.

Reder, L. M., and Anderson, J. R. (1980). "A comparison of texts and their summaries: Memorial consequences", *Journal of Verbal Learning and Verbal Behavior*, 19(2), pp. 121-134.

Royer, J. (1990). "The sentence verification technique: a new direction in the assessment of reading comprehension", in Legg, S. and Algina, J. (eds.), *Cognitive Assessment of Language and Math Outcomes*, Norwood, NJ: Ablex.

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

Shohamy, E., and Inbar, O. (1991). "Validation of listening comprehension tests: The effect of text and question type", *Language Testing*, 8(1), pp. 23-40.

van den Broek, P., Fletcher, C. R., and Risden, K. (1993). "Investigation of inferential processes in reading: A theoretical and methodological integration", *Discourse Processes*, 16, pp. 169-180.

van den Broek, P., and Kremer, K. (1999). "The mind in action: What it means to comprehend during reading", in Taylor, B. M., van den Broek, P., and Graves, M. (eds.), *Reading for Meaning*, New York: Teachers College Press.

Vongpumivitch, V. (2004). Measuring the knowledge of text structure in academic English as a second language (ESL). Unpublished PhD dissertation, UCLA.

Weir, C. (1993). *Understanding and Developing Language Tests*, London: Prentice Hall.

Williams, J. P. (2005). "Instruction in reading comprehension for primary-grade students: a focus on text structure", *Journal of Special Education*, 9(1), pp. 6-18.

Williams, J. P., Brook Stafford, K., Lauer, K. D., Hall, K. M., and Pollini, S. (2009). "Embedding reading comprehension training in content-area instruction", *Journal of Educational Psychology*, 101(1), pp. 1-20.