

«نشریه علمی-پژوهشی آموزش و ارزشیابی»

سال هشتم - شماره ۳۲ - زمستان ۱۳۹۴

ص. ص. ۱۵۶-۱۳۵

The Effect of Text Structure on Iranian EFL Learners' Reading Comprehension as Measured by Multiple-choice and Cloze Tests

Hedayat Eslami¹

Mahnaz Saeidi^{2*}

Saeide Ahangari³

تاریخ دریافت مقاله: ۱۳۹۴/۰۸/۳۷

تاریخ پذیرش نهایی مقاله: ۱۳۹۴/۱۲/۰۱

Abstract

This study was an attempt to investigate Iranian EFL students' performance on reading comprehension tests with different text structures and response formats. The participants of the study included 228 students, comprising eight groups, selected based on the Preliminary English Test (PET). Four types of text structures, time sequence (T), description (D), causation (C), problem-solution (P) were selected and two types of response formats, Multiple-Choice Questions (MCQ) and Multiple-Choice Cloze tests (MCC), based on the four text structures were developed and administered to the eight groups. The results revealed that in both response formats, the students' performance on the more loosely organized text structures (T and D) was better than their performance on the more tightly organized ones (C and P). Furthermore, the students' performance on MCQ was significantly better than on MCC response format across all four text structures. The results of the study suggested that students performed better on T and D text structures compared to C and P in both response formats and that their performance on MCQ response format was better than their performance on MCC across all four text structures. The results of the study suggest that teachers and test developers can boost their students' performance on reading comprehension tests by choosing the most appropriate response formats for different text structures.

Keywords: text structure, response formats, multiple-choice questions, multiple-choice cloze

1 -2 &3 Department of English, Tabriz branch, Islamic Azad University, Tabriz, Iran (m_saeidi@iaut.ac.ir)

2* Corresponding Author

تأثیر ساختار متن بر عملکرد خواندن و درک مطلب زبان آموزان ایرانی در آزمون‌های چهار جوابی و کلوز

هدایت اسلامی

مهناز سعیدی

سعیده آهنگری

چکیده

این تحقیق به منظور بررسی عملکرد زبان آموزان ایرانی در آزمون‌های خواندن و درک مطلب دارای ساختارهای متنی و روش‌های پاسخ‌دهی متفاوت انجام گرفته است. شرکت‌کنندگان این تحقیق ۲۲۸ نفر از دانشجویان زبان بودند که شامل ۸ گروه و بر اساس آزمون مهارت زبانی پت انتخاب شدند. چهار نوع ساختار متنی ترتیب زمانی، توصیفی، علیت و حل مسئله انتخاب و بر اساس این متون، دو نوع آزمون با شیوه پاسخ‌دهی چهارجوابی و چهارجوابی کلوز تهیه و به دانشجویان داده شد. نتایج بدست آمده نشان داد که در هر دو روش پاسخ‌دهی عملکرد دانشجویان در متون دارای انسجام متنی ساده (ترتیب زمانی و توصیفی) بهتر از عملکرد آنان در متون دارای انسجام متنی پیچیده (علیت و حل مسئله) است. علاوه بر این، عملکرد دانشجویان در روش پاسخ‌دهی چهارجوابی بطور معنی‌دار بهتر از عملکرد آنها در روش پاسخ‌دهی چهارجوابی کلوز در تمامی چهار نوع ساختار متنی بود. در پایان چنین نتیجه‌گیری شد که دانشجویان در ساختارهای متنی ترتیب زمانی و توصیفی عملکرد بهتری نسبت به ساختارهای متنی علیت و حل مسئله در هر دو روش پاسخ‌دهی دارند و همچنین در روش پاسخ‌دهی چهارجوابی بهتر از روش پاسخ‌دهی چهارجوابی کلوز در هر چهار نوع ساختار متنی هستند. معلمان می‌توانند با بهره‌گیری از نتایج این تحقیق و از طریق انتخاب بهترین روش پاسخ‌دهی برای متون دارای ساختار متنی متفاوت عملکرد دانشجویان خود را در خواندن و درک مطلب بهبود بخشند.

واژگان کلیدی: ساختار متن، روش پاسخ‌دهی، چهارجوابی، کلوز

Introduction

The investigation of test method effect is an ongoing tradition in testing research that is often conducted in the area of reading tests (Alderson & Banerjee, 2002). It is argued that performance on language test is affected by the characteristics of the methods used to elicit test performance (i.e., test method facets). These characteristics, or facets, of test methods constitute the how of language testing and are of great significance in designing, developing, and using language tests (Bachman, 1990). Rupp, Ferne, and Choi (2006) also argued that the design of items and the selection of texts

on a reading comprehension assessment are major factors which should be taken into consideration because they operationalize the complex construct of reading comprehension and the process of making sense of printed text in non-testing contexts. Bachman (1990) and Bachman and Palmer (1996) developed a theory of language testing which encompassed not only different aspects of language ability but also the methods and other factors involved in the measurement of this ability. Bachman's (1990) theoretical framework of test method facets consists of five major categories including a) the testing environment, b) the test rubric, c) the nature of input, d) the nature of expected response, and e) the interaction between input and response.

Following Bachman's (1990) model of test method facets, a plethora of studies (e.g., Atai & Soleimany, 2009; Cheng, BiglarBaygi, & Solaymani, 2009; Kobayashi, 2002, 2004; Liu, 2009; Ozuru, Best, Witherspoon, & McNamara, 2007; Rahimi, 2007; Rauch & Hartig, 2010; Sadighi, Yamini, & Ayatollahi, 2007; Shahivand & Pazhakh, 2012; Shohamy, 1984; Tavakoli, Ahmadi, & Bahrani., 2011; Wolf, 1993) have demonstrated that the methods used to measure language ability influence performance on language tests verifying Bachman's (1990) assertion concerning the effect of test method facets on language test performance. It follows that language testers should take stock of the testing methods because they can influence performance on language tests and therefore jeopardize test validity (Bachman, 1990). A number of studies have been undertaken concerning the effect of different facets of test method based on Bachman's model (1990) among which the investigation of text structure and response format have received utmost attention. These two facets to be investigated in this study are related to the nature of input and the nature of expected response in Bachman's (1990) test method model.

The structure of text, as an aspect of the nature of input, refers to the material presented to the test-takers and is deemed to play a vital role in enhancing the efficacy of the reading tests, and can influence test performance (Bachman, 1990). Text structure, or text organization is the way the paragraphs relate to each other and the way the relationships between ideas are signaled or not signaled (Alderson, 2000). Text structure is inherent in a text's organizational pattern, which reflects the logical connections among the ideas in the text (Meyer & Poon, 2001). That is why, texts are better understood through the readers' interpretation of the larger organizational structure signaled by the writer (Grabe, 2002). In line with these assertions, some studies (e.g., Bachman, 1990; Fountas & Pinell, 2001; Jiang, 2012; Kobayashi, 2002, 2004; Meyer & Poon, 2001; Paltridge,

1996; Sadighi et al., 2007; Sharp, 2002, 2004; Williams et al., 2014; Zhang, 2008) have endorsed the effect of text organization and text characteristics on students' performance on reading comprehension tests. As a result, language testers may get different pictures of comprehension depending on the combination of assessment tools, age groups, genre of the text used in the assessment, and text characteristics within a genre (Magliano, Millis, Ozuru, & McNamara, 2007).

The effect of response formats, as an aspect of the nature of the expected response in Bachman's (1990) model, is also deemed to be a crucial factor affecting performance on language tests (Alderson, 2000; Bachman, 1990; Bachman & Palmer, 1996; Brantmeier, 2005; Chehrazad & Ajideh, 2013; Coombe, Folse, & Hubley, 2007). Therefore, language teachers should be wary of choosing different test formats because the type of response can impact the students' ability on a language test (Coombe et al., 2007). For example, a student may be able to select a correct answer when asked about the plot or sequence of a story, but the same student may have difficulty in supplying or creating an answer which indicates that he actually understands the content of the story. Consequently, in the recent years, there has been an increase in the number of different techniques used for testing reading comprehension which range from different objective techniques such as multiple-choice to non-objective methods like short-answer, or even summaries which have to be subjectively evaluated (Alderson, 2000). Tests of MCQs, cloze, gap-filling, editing, matching, recall, summary, cloze-elide, short-answer are a few examples in this regard (Alderson, 2000; Brantmeier, 2005; Brown, 2004).

The empirical studies abound in the area of test method facets and their effects on language test performance, however their scopes were limited. Liu (1998) and Liu (2009) focused just on the effect of three test methods on performance on reading comprehension tests without considering text types. Using a multiple-choice, true-false, and short-answer questions in his reading comprehension test, Liu (1998) found that there were significant differences among the scores of the three groups and that short answer questions were the most difficult. By the same token, Liu (2009) scrutinized the effect of three test methods, namely, MCQs, gap-filling, and short-answer questions on reading comprehension. The study revealed that gap-filling test was the most difficult while MCQs and short-answer questions were the easiest.

The study by Ajideh and Esfandiari (2009) focused on comparing the effect of two test formats of multiple-choice test of lexical items and cloze test with fixed-ratio deletion. They concluded that in testing the

proficiency of a group of learners, the achieved scores on multiple-choice test were much similar to the cloze test scores. Although the two tests were seemingly different, there was a high correlation between the two types of test formats on vocabulary discrete-point items and integrative cloze test. Unlike Ajideh and Esfandiari's (2009) study, Shahivand & Pazhakh (2012) demonstrated that the cloze test was the most difficult form of testing compared to MCQs and true-false items. They also found that the discrete-point items were easy to answer because they measure one aspect of the language and students could answer the items more easily than integrative test items. Likewise, Chehrazad and Ajideh (2013) examined the effect of MCC and MCQs on pre-intermediate and intermediate test-takers' reading comprehension performance. They found no significant differences between the two groups' performance on MCC test, however, they demonstrated that the intermediate test-takers significantly outperformed the pre-intermediate ones on MCQs.

In a different study, Rahimi (2007) investigated the effect of presenting the items of an English reading comprehension test in the testees' native language (Persian) on their performance on the test. He indicated that test method (i.e., the language of presentation) did not significantly affect the students' scores on a reading comprehension test' on the whole, however, the test method was found to affect the performance of low proficient testees.

There are three studies which involved probing the effect of both text structure and response formats on EFL students' performance on reading different types of texts, however each of which focused on a particular response format. To cite as an example, Kobayashi (2002, 2004) conducted a series of studies on the effect of text organization and response format on L2 learners' performance on reading comprehension tests. She found that text structure affected students' performance on a reading test and their performance on the more loosely organized texts such as association (i.e., T) was better than their performance on the more tightly organized ones like C and P when their comprehension was measured by cloze test. She also demonstrated that the students' performance was affected by the type of response format used, that is, their performance on the more tightly organized texts was better on open-ended and summary writing, but worse on a cloze test. Based on Meyer's (1975) model of text structure, Sadighi et al. (2007) found significant differences in the performance of participants on the four types of text structures, namely, collection, D, C, and P as measured by structured and unstructured summary writing. They found that their participants had the highest scores in the collection type of text (a more

loosely organized text) followed by the C text structure (a more tightly organized text) both in structured and unstructured summary writing. In line with these studies, Akhondi and Malayer (2010) compared the effectiveness of three response formats (i.e., incomplete outline, graphic organizers, summary writing) to gauge TESL students' knowledge of text structure in Malaysia. They found that high-achievers outperformed intermediate-and low-achievers across the three response formats. Furthermore, the three groups achieved higher scores on incomplete outline and summary writing, respectively. Nonetheless, the graphic organizer appeared as the most difficult task since the respondents achieved the lowest score in this task.

Drawing upon the pertinent literature and considering subsequent empirical studies, one can argue that the structural features of texts (e.g., causal and rhetorical structures) are important factors to be included in the assessment tools since these structural features affect the degree to which readers engage in strategic processing of texts (Trabasso et al., 1984, as cited in Magliano et al., 2007). Moreover, the type of response formats have been shown to influence performance on language tests (e.g., Atai & Soleimany, 2009; Bachman & Palmer, 1996; Bachman, 1990; Cheng et al., 2009; Chehrazad & Ajideh, 2013; Kobayashi, 2002, 2004; Liu, 2009; Rauch & Hartig, 2010; Shahivand & Pazhakh, 2012; Shohamy, 1984). Language teachers should, therefore, be cognizant of the effect of these variables as facets of test methods and strive to minimize their influence because they can jeopardize the validity of the tests (Bachman, 1990).

As there is dearth of research studies regarding four types of text structures (i.e., T, D, C, P) and two types of response formats (i.e., MCQ and MCC), the present study attempted to contribute to research literature by focusing on these particular text structures and response formats. Some of the earlier studies conducted so far have focused only on the effect of response formats (e.g., Ajideh & Esfandiari, 2009; Chehrazad & Ajideh, 2013; Cheng et al., 2009; Salmani Nodoushan, 2010; Shahivand & Pazhakh, 2012), or on the effect of text structure through other response formats rather than MCQ and MCC; for example, structured and unstructured summary (Sadighi et al., 2007); still others have manipulated other features of texts (as facets of test methods) like the effect of genre familiarity via cloze and C-test (Tavakoli et al., 2011) and text authenticity and genre through C-test (Atai & Soleimany, 2008; Cheng et al., 2009).

As a result, this study focusing on the third and fourth facets of Bachman's (1990) theoretical model of test method facets, namely, the nature of the input (i.e., text structure) and the nature of expected response (i.e., types of response formats) investigated the effect of text structure (i.e.,

T, D, C, P) and response formats (i.e., MCQs and MCC) on Iranian intermediate EFL learners' performance on reading comprehension tests.

The current study was designed to answer the following three research questions regarding the effect of text structure and response format on EFL students' performance on reading comprehension tests:

RQ 1: Are there any significant differences in the performance of Iranian intermediate EFL learners on four types of text structures (i.e., T, D, C, P) across MCQ and MCC response formats in reading comprehension tests?

RQ 2: Are there any significant differences between intermediate EFL learners' performance on MCC and MCQ response formats across four types of text structure (i.e., T, D, C, P) in reading comprehension tests?

RQ 3: Are there any main effects or interaction effect between Iranian EFL learners' performance on four types of text structure (i.e., T, D, C, P), and two types of response format (i.e., MCQ and MCC) in reading comprehension tests.

Method

Participants

The participants of the study included 448 Iranian EFL students majoring in Teaching English as a Foreign Language (TEFL), English Literature, and English Language Translation (ELT) from Islamic Azad University, Tabriz branch and Payame-Noor University, Miyndoab Center. They were junior and senior male and female EFL students with the age range of 21-30 and their language background was Azeri, Kurdish, and Farsi. To accomplish the purpose of the study, 448 students took part in the pilot study and the main study. Two hundred and twenty of these students participated in the pilot study, which took place in two phases (phase one = 41, and phase two = 179) and 228 students, comprising eight groups, participated in the main study. The proficiency level of the participants in the main study was checked by Preliminary English Test (PET).

3.3. Instruments

The first instrument used was Preliminary English Test (PET). The reading section comprised five parts with 35 questions and the writing section consisted of three parts with 7 questions.

The second instrument was a set of reading comprehension tests developed in the two response formats of MCC and MCQs and based on four types of texts with different structures (i.e., T, D, C, P). The MCC and MCQs were piloted in two phases in order to ascertain their psychometric characteristics. First, the open-ended form of MCC was piloted with a sample of 18 and 23 EFL students with characteristics similar to the target

groups in order to elicit efficient distracters from the participants. Then, in the second phase of piloting, the four types of MCQs and four types of MCC tests were piloted with a number of 152 other students and their reliability was estimated (i.e., MCQT=.63, MCQD= .66, MCQC= .60, MCQP= .74, MCCT= .83, MCCD = .88, MCCC = .87, M CCP= .89). The final versions of MCQ and MCC tests comprised 20 questions and 40 blanks, respectively.

Procedure

The study took place in two stages of pilot study and main study:

Pilot study

In the pilot study, first a number of reading passages with four different types of text structures (i.e., T, D, C, P) were selected and matched to the EFL students' ability level using Flesch-Kincaid Reading Ease Score. In the selection of texts, the content of the passage, the definition of and explanation for each text type, and the use of signals words in each text structure were taken into consideration in order to determine the appropriate text in terms of the organizational patterns. Then, a few experts with Ph. D degree in TEFL were asked to review these texts in order to ensure the content validity of the material in terms of each text type. In the next step, based on these text structures, eight types of reading comprehension tests in the two response formats of MCQ and MCC were developed and piloted in two phases in order to ascertain their psychometric characteristics.

Main Study

In the main study, two necessary conditions had to be met: the homogeneity of the eight intact classes and the normality of the distribution of scores of the eight groups in the eight reading tests. To this end, first, PET was administered to ensure that the eight intact classes were homogeneous in terms of language proficiency. Therefore, a number of 228 students participated in this phase of the study (i.e., G1=28; G 2= 30; G 3= 26; G 4= 30; G 5= 32; G 6= 28; G 7= 28; G 8= 26).

In the next step, the eight newly developed reading comprehension tests in the two response formats of MCQ and MCC were administered to eight intact homogeneous classes. Therefore, four groups took the MCQ tests of T (MCQT), of D (MCQD), of C (MCQC) and of P (MCQP) and four groups took the MCC tests of T (MCCT), of D (MCCD), of C (MCCC) and of P (MCCP). The eight sets of tests were arranged so that each version was randomly distributed among the eight participant groups. Having accomplished this, Kolmogorov-Smirnov (K-S) method was utilized in

order to ascertain the normality of the distributions of scores and determine the type of appropriate statistical tests (i.e., parametric versus non-parametric tests). After administering the eight reading comprehension tests, the students' answers on these tests were collected for data analysis.

Results

The Results of Proficiency Test

The results of one-way ANOVA, as shown in Table 1, run on the scores of eight participant groups on the reading and writing sections of language proficiency test attested the homogeneity of the 8 groups in terms of language ability as the differences among their mean scores were not statistically significant ($p > .05$).

Table 1. The results of one-way ANOVA for the homogeneity of 8 groups on the proficiency test

	N	Mean	Std. Deviation	Minimum	Maximum	ANOVA	
						F	P value
MCQT	32	32.7436	13.75102	10	54	.192	0.967
MCQD	28	29.8718	14.23520	10	52		
MCQC	28	30.7674	14.05573	10	54		
MCQP	30	31.8537	14.22596	10	54		
MCCT	26	32.4286	13.11514	10	54		
MCCD	30	31.6667	12.79227	10	54		
MCCC	28	31.9524	13.99120	10	53		
MCCP	26	30.8462	13.80797	10	54		

The Results of Reading Tests

Before analyzing the students' scores on the 8 reading tests, the normality of the distribution of the groups' scores on these tests was ascertained using one-sample Kolmogorov-Smirnov test the results of which indicated normal distribution of scores for all groups since the p value exceeded .05 (see Table 2).

Table 2. One-Sample Kolmogorov-Smirnov test for the normality of the distribution of scores in the 8 groups

	N	Kolmogorov-Smirnov Z	P value
MCQT	32	0.791	0.558
MCQD	28	0.820	0.512
MCQC	28	0.755	0.619
MCQP	30	0.941	0.339
MCCT	26	0.578	0.892
MCCD	30	0.644	0.801
MCCC	28	0.629	0.823
MCCP	26	0.823	0.508

After ensuring the normality of the distribution of scores in the 8 reading test, the students' scores on these tests (i.e., MCQT, MCQD, MCQC, MCQP, MCCT, MCCD, MCCC, & MCCP) were analyzed.

Testing the First Research Hypothesis

The first hypothesis concerning the effect of text structure stated that there were significant differences in the performance of Iranian intermediate EFL learners on reading comprehension tests with four types of text structure (i.e., T, D, C, P) across MCQ and MCC response formats.

In order to probe the differences in the text type effects on the participants' performance on reading comprehension tests, their mean scores on four types of text structure were compared in each response format. These results are illustratively presented in Figure 1.

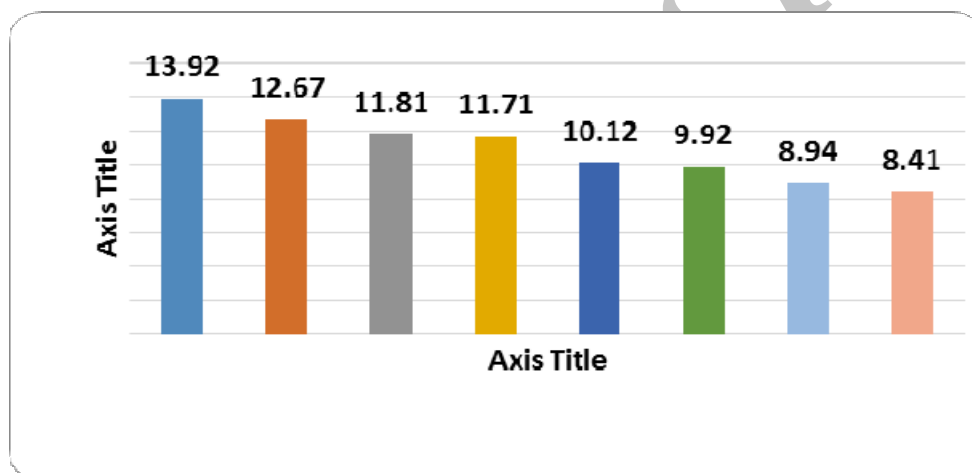


Figure 1. The comparison of performance on four types of text structures in each response format.

Figure 1 illustrates that the mean scores of four text types in each response format are different from each other. The results of two one-way ANOVAs conducted on the mean scores of the four groups in each response format indicated that the differences were statistically significant as the P value was smaller than .05 ($p < .05$) (see Table 3). Therefore, the null hypothesis of no significant difference among the performance of the students on four types of text structures can be safely rejected, and the alternative hypothesis stating that there were significant differences in the performance of Iranian intermediate EFL learners on reading comprehension tests with four types of text structure (i.e., T, D, C, P) across MCQ and MCC response formats was confirmed.

Table 3. The results of one-way ANOVA for the four types of texts in each response format

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		ANOVA	
						Lower Bound	Upper Bound	F	P value
MCQ	Time	32	13.92	2.71	.43	13.05	14.80	3.404	.019
	Description	28	12.67	3.61	.58	11.50	13.84		
	Causation	28	11.81	3.76	.57	10.66	12.97		
	Problem-S	30	11.71	3.82	.60	10.50	12.91		
MCC	Time	26	10.12	1.34	.64	8.82	11.41	4.67	.012
	Description	30	9.92	1.42	.59	8.72	11.11		
	Causation	28	8.94	1.37	.63	7.67	10.22		
	Problem-S	26	8.41	1.11	.73	6.93	9.89		

In order to locate the places of differences among the four text types in each response format, use was made of a post-hoc analysis, multiple Scheffee test, for pairwise comparison of the performance of the four groups in each response format (see Table 4).

Table 4. Multiple comparisons Scheffe test for four types of text in MCQ and MCC response formats.

(I)	(J)	95% Confidence Interval			
		MCQ		MCC	
		Mean Difference (I-J)	P value	Mean Difference (I-J)	P value
	description	1.26	0.479	0.2	0.231
time	causation	2.11	0.045	1.18	0.041
	problem-s	2.22	0.041	1.71	0.035
description	causation	0.85	0.752	0.98	0.047
	problem-s	0.96	0.685	1.51	0.039
causation	problem-s	0.11	0.999	0.53	0.124

The results of this post hoc analysis for pairwise comparison in MCQ test format indicate that: a) the mean score of MCQT is not significantly different from the mean score of MCQD ($p > .05$) implying no statistically significant difference in the performance of the participants between these two types of texts. However, the mean score of MCQT is significantly different from the mean scores of MCQC and MCQP ($p < .05$) implying that the performance of participants in MCQT test was better than their performance on MCQC and MCQP groups; b) the mean score of MCQD is not significantly different from the three other tests ($p > .05$); c) the mean scores of MCQC and MCQP are not significantly different from each other.

With respect to MCC, the results of this post hoc analysis for pairwise comparison in MCC test format indicate that: a) the mean score of MCCT is not significantly different from the mean score of MCCD ($P > .05$), however, it is significantly different from the mean scores of MCCC and MCCP ($P < .05$); b) the mean score of MCCD is significantly different from the mean scores of MCCC and MCCP ($P < .05$); c) the means scores of MCCC and MCCP are not significantly different from each other ($P > .05$).

Testing the Second Research Hypothesis

The second research hypothesis with regard to the effect of response format stated that there were significant differences between intermediate EFL learners' performance on MCQ and MCC response formats across four types of text structures (i.e., T, D, C, P) in reading comprehension tests.

In order to probe the differences on MCQ and MCC response formats across all four types of text structures, the scores of the participants on these two response formats across four text structures were compared. The mean scores of the students on two response formats for each text type are illustrated in Figure 2.

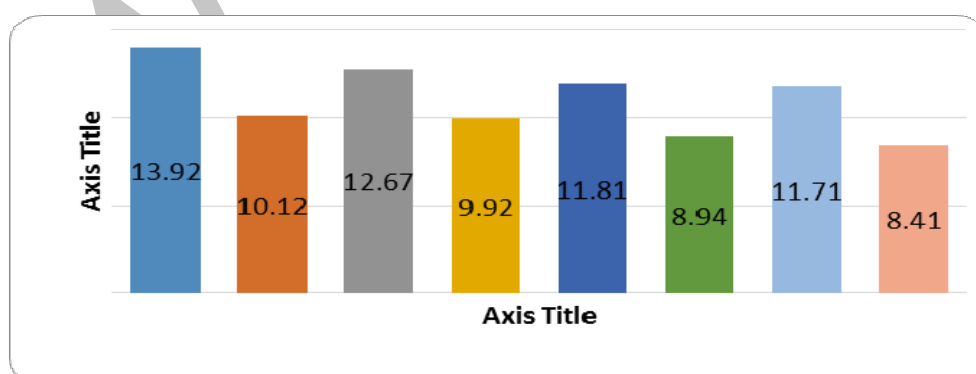


Figure 2. The comparison of performance in the two response formats across four text types

To test the mean scores of the two response formats in each reading test for statistical significance, four Independent Samples T-tests were conducted the results of which are illustrated in Table 5.

Table 5. The results of Independent Samples T-tests comparing the performance of two response formats in each of the four text type.

Text Organization	Response Formation	N	Mean	Std.	t-test for Equality of Means			
					t	df	p value	Mean Difference
Time sequence	MCQ	32	13.92	2.71	4.83	56	0.000	3.8
	MCC	26	10.12	1.34				
Description	MCQ	28	12.67	3.61	3.32	56	.001	2.75
	MCC	30	9.92	1.42				
Causation	MCQ	28	11.81	3.76	3.37	54	.001	2.87
	MCC	28	8.94	1.37				
Problem-S	MCQ	30	11.71	3.82	3.51	54	.001	3.29
	MCC	26	8.41	1.11				

According to Table 5, the results of p value for the difference between the mean scores of the two groups (MCQ vs. MCC) across all four text types is statistically significant as the p value is smaller than .05 ($P < .05$). Therefore, the null hypothesis of no significant difference between EFL students' performance on MCC and MCQ across all four text types can be safely rejected, and the alternative hypothesis stating that there were significant differences between intermediate EFL learners' performance on MCQ and MCC response formats across four types of text structures (i.e., T, D, C, P) in reading comprehension tests was confirmed.

Testing the Third Research Hypothesis

The third research hypothesis was concerned with examining the main effects and interaction effect between Iranian EFL learners' performance on four types of text structure (i.e., T, D, C, P), and two types of response format (i.e., MCQ and MCC) in reading comprehension tests.

In order to test this hypothesis, the researchers employed a two-way ANOVA to assess the main effects and the interaction effect (i.e., response formats and text structures) the results of which are presented in Table 6.

Table 6. The results of two-way ANOVA to assess the main effects and interaction effect

Source	Type III sum of square	df	Mean Square	p value
Response	838.414	1	838.414	.001
Text	129.833	1	43.278	.002
Response*text	3.522	3	1.174	.853
Error	614.014	56	4.489	
Total	44721.000	228		

According to Table 6, the p value for the main effects of response format and text structure are statistically significant ($p < .05$). However, the interaction effect between these two variables ($p = .853$) is not statistically significant ($p > .05$).

With regard to the main effect of the response formats, first the descriptive statistics were computed for the MCQ and MCC (Table 7).

Table 7. The descriptive statistics for the two response formats.

group	N	Mean	Std. Deviation	Std. Error Mean
MCQ	118	12.5000	3.59131	.28216
MCC	110	9.3636	4.17979	.32540

Second, an independent Samples t-test was utilized to compare their mean scores for statistical significance, the results of which are presented in Table 8.

Table 8. Independent Samples T-test results for the main effect of response formats.

t	df	P value	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
7.272	226	.000	3.13636	2.28789	3.98484

The results of Table 8 indicate that the difference in the mean scores between MCQ and MCC is statistically significant ($P < .05$). This means that

in MCQ, the participants significantly outperformed those in MCC response format.

Regarding the main effect for text structure, likewise, the descriptive statistics for the four types of text structures were computed the results of which are indicated in Table 9.

Table 9. The mean and Std. D of scores in the four types of text structures.

group	N	Mean	Std. Deviation	95% Confidence Interval for Mean	
				Lower Bound	Upper Bound
Time	58	11.9506	4.00047	11.0660	12.8352
Description	58	11.2407	3.94898	10.3675	12.1139
Causation	56	10.3941	4.16125	9.4966	11.2917
Problem-S	56	10.1000	4.48866	9.1011	11.0989

According to Table 9, there are differences in the mean scores of four types of text structure. In order to ensure for the significance of differences among the mean scores of the four groups, a one-way ANOVA was utilized the results of which are provided in Table 10.

Table 10. The results of One-way ANOVA for the main effect of text structure.

	Sum of Squares	df	Mean Square	F	P value
Between Groups	171.666	3	57.222	3.316	.020
Within Groups	5574.105	225	17.257		
Total	5745.771	228			

The results of this analysis, as shown in Table 10, illustrates that there are significant differences among the mean scores of the four groups ($P < .05$). To locate the places of differences, a post hoc analysis, Scheffee Test, was run for pairwise comparisons, the results of which are shown in Table 11.

Table 11. The results of Scheffee test for pairwise comparisons of the mean scores of four types of text structures.

(I)	(J)	Mean Difference (I-J)	P value	95% Confidence Interval	
				Lower Bound	Upper Bound
Time	Description	.70988	.757	-1.1246	2.5443
Time	Causation	1.55650	.123	-.2563	3.3693
Time	Problem-S	1.85062*	.048	.0104	3.6908
Description	Causation	.84662	.632	-.9661	2.6594
Description	Problem-S	1.14074	.388	-.6994	2.9809
Causation	Problem-S	.29412	.976	-1.5244	2.1127

The results of Table 11 for the main effect of text structure indicate that the difference lies in T text structure because the participants' performance on this type of text is significantly higher than the performance of participants in P text structure ($P < .05$). However, it is not significantly different from the performance on D and C types of text structures.

Discussion and Conclusion

The present study was an attempt to scrutinize the effect of four types of text structures, time sequence (T), description (D), causation (C), and problem-solution (P) on performance in reading comprehension tests across MCQ and MCC response formats.

The results of the analysis of students' performance on four text structures in MCQ response format indicated that students in MCQT test significantly outperformed those in MCQC and MCQP tests. This suggests that students' performance on a more loosely organized text like T was better than their performance on more tightly organized texts such as C and P when measured via MCQ. In MCC response format, the results indicated that students in MCCT and MCCD text structures significantly outperformed those in MCCC and MCCP tests. This again suggests that students can perform better on more loosely organized texts and worse on more tightly organized ones. It follows that the structure of text is more effective and facilitative in reading comprehension when the texts are more loosely organized as in T and/or D and when they are gauged via MCQ and MCC. A possible explanation for the students' better performance on the loosely organized text structures is that in these types of texts, the relationships among ideas are not complex or that these texts have simple text structures that make the comprehension processing easier. These results are in line with studies conducted previously on the effect of text structure on performance in reading comprehension test as measured by cloze test (e.g., Fountas & Pinell, 2001; Kendeou & Broek, 2007; Kobayashi, 2002; Sharp, 2002, 2004; Yoshida, 2012). On the other hand, the students' weak performance on more tightly organized texts may be due to the complex organizational patterns inherent in these types of texts, especially in C text structure, or it may be due to the limits of the learners' short-term memory to process the conceptual relationships in the more complex organizational patterns, such as C and P, as Alderson (2000), Sharp (2002), and Snyder (2012) asserted.

More importantly, the results of this study contradict Meyer (1975, 1985) and Carrell (1985) who claimed that the presence of clear text structures in texts like C and P facilitate reading comprehension. However,

the results of present study showed that the more tightly organized texts (e.g., C) were not conducive in facilitating performance on reading comprehension implying that Meyer's (1975, 1985) claim concerning the facilitative role of more tightly organized texts in recall cannot be extended to measuring reading comprehension through MCQ and MCC. In this sense, the results of this study corroborate and are in line with studies carried out by Kobayashi (2002), Sharp (2002), and Sadighi et al. (2007) who came up with similar results on the effect of loosely organized texts on facilitating reading comprehension. Kobayashi (2002; 2004), for instance, demonstrated that students perform better on more loosely organized texts when their comprehension is measured by cloze test. Sharp (2002) also found that students obtained significantly higher scores with more loosely organized texts (i.e., D) and the lower scores with more tightly organized texts (i.e. C) when gauged via cloze procedure. More interestingly, Sharp (2002) found no significant difference in the performance of the learners on the four rhetorically different texts as measured via recall, again contradicting the results obtained by Meyer (1975, 1985) and Carrell (1985). Sadighi et al. (2007) came up with similar results concerning the facilitative role of more loosely organized texts in reading comprehension as measured via structured and unstructured summary writing. They found that students had the highest scores in collection type of text which is a more loosely organized text compared to the C text structure which is a more tightly organized text.

With respect to the effect of response formats on EFL students' performance on reading comprehension, the results of four independent samples T-tests on MCQ and MCC response formats in each of the four text types revealed that students in MCQ consistently outperformed those in MCC test format, and this outperformance was statistically significant. This means that different test methods (response formats) affect students' performance differently in reading comprehension tests.

According to the results, the MCQ response format was easier than MCC format across all four types of text structures. In this sense, the results of this study contradict Kobayashi's results on the effect of response format on C text because Kobayashi (2002) found that the effect of response format was not statistically significant for this type of text while it was statistically significant for association (i.e., T), D, and P. However, the results of the present study indicated that the effect of two response formats was statistically significant across all four text structures. One probable reason for the students' better performance on MCQ test may be ascribed to their familiarity with these types of tests, which is the result of their widespread

previous encounters with these test formats. Thus, their weak performance on MCC might be due to the fact that cloze tests are not as prevalent as MCQ tests in testing students' reading comprehension, so lack of the students' familiarity with the cloze test might contribute to the obtained results.

In this sense, the results of this study are consistent with Chehrazad & Ajideh (2013), Liu (1998, 2009), Samson (1983), Shahivand and Pazhakh (2012), Shohamy (1984), Wolf (1993) all of whom found that MCQ format was the easiest test format among the cloze test, gap-filling, short answer, open-ended, and summary writing. Chehrazad & Ajideh (2013), for example, indicated that intermediate test-takers' performance on MCQ test was significantly better than their performance on MCC test on a test of reading comprehension. The results of this study, however, are at odds with the studies by Ajideh and Esfandiari (2009) who found that students' scores on multiple-choice test of lexical items were much similar to the integrative cloze test, and Sun (2001) who found no significant differences among the three test methods of multiple-choice, true/false, and short answers on Grade two Junior Middle school students in China. The students' high performance on MCQ in comparison to their performance on other types of response formats may imply that the two formats are measuring different skills or different aspects of the same skill. Several empirical studies accord with this suggestion (e.g., Kobayashi, 2002; Rupp et al., 2006; Shohamy, 1984). Shohamy (1984), for instance, argued that the learners' high performance on MCQ suggests that a different skill is required in order to fulfill the MCQ task. Therefore, in conformity with the pertinent literature, it seems that response formats can induce variations in language test performance and affect it. In this respect, Kobayashi (2002) argued that variation in the students' performance on different types of response formats may suggest that different test formats seem to gauge different aspects of reading comprehension.

Finally, the analysis of results revealed that the main effects for both response formats and text structures are statistically significant ($p < .05$). However, there was no interaction effect between these two variables. In this sense, the results contradict Kobayashi's (2002) study in which she found an interaction between text structure and response formats.

In conclusion, in the light of Bachman's (1990) theoretical model of test method facets, the study examined the effect of text structure and response formats on Iranian EFL learners' performance in reading comprehension tests. Regarding the effect of text structure on reading comprehension, the results revealed that the EFL students' performance on

the more loosely organized text structures (i.e., T and D) was significantly better than their performance on the more tightly organized ones (i.e., C and P) across the two response formats of MCQ and MCC. With respect to the effect of response format, the results showed that in MCQ response format, learners' performance was significantly better than their performance on MCC format across all four text structures. Therefore, it can be concluded that text structure and response format significantly affect EFL learners' performance in reading comprehension tests verifying the previous assertions made in this regard (e.g., Atai & Soleimany, 2009; Bachman, 1990; Bachman & Palmer, 1996; Cheng et al., 2009; Kobayashi, 2002, 2004; Liu, 1998, 2009; Ozuru et al., 2007; Rahimi, 2007; Sadighi et al., 2007; Shahivand & Pazhakh, 2012; Shohamy, 1984; Tavakoli et al., 2011; Wolf, 1993).

EFL teachers can use the results of this study on text structure to enhance their students' reading comprehension performance and test developers will be able to minimize the effect of text structure and response formats, as the intervening factors, in designing their language tests. They will also be able to choose and use the most appropriate response formats in measuring their students' performance in reading comprehension tests.

Considering the limitations of the present study, it should be pointed out that due to practical problems, the study excluded a number of other response formats and just focused on MCQ and MCC, and due to time limitation, the researchers had to include only two reading passages for each reading test. Researchers can scrutinize the effect of other types of response formats such as open-ended cloze test, summary writing, short-answer, true/false on ESP or EFL students' performance on reading comprehension tests. It would also be interesting to examine the effect of discussion of text structure on students' reading comprehension based on the sociocultural framework. Investigating the effect of other factors such as test takers' characteristics, test rubrics, item formats, and background knowledge on EFL students' performance on reading comprehension tests are other areas of interest for those interested in examining factors affecting the EFL students' performance on reading comprehension tests.

References

- Ajideh, P., & Esfandiari, R. (2009). A close look at the relationship between multiple-choice vocabulary test and integrative cloze test of lexical words in Iranian context. *English Language Teaching*, 2 (3), 163-170.
- Ajideh, P., & Mozaffarzadeh, S. (2012). A comparative study on C-test vs. cloze test as tests of reading comprehension. *Journal of Basic and Applied Scientific Research*, 2 (11), 1159-11163.
- Akhondi, M. & Malayeri, F.A. (2010). Assessing reading comprehension of expository text across different response formats. *Iranian Journal of Applied Studies*, 3 (1), 1-26.
- Alderson, J.C. (2000). *Assessing reading*, Cambridge: Cambridge University Press.
- Alderson, J.C., & Banerjee, J. (2002). Language testing and assessment. *Language Teaching*, 35, 79-113. Doi: 10.1017/S0261444802001751.
- Atai, M.R., & Soleimany, M. (2009). On the effect of text authenticity and genre on EFL learners' performance in C-tests. *Pazhuhesh-e Zabanha-ye Khareji*, 49, 109-123.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal*, 89 (1), 37-53.
- Brown, H.D. (2004). *Language assessment: Principles and classroom practices*. U.S.A: Longman.
- Carrell, P. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19, (4), 727-752.
- Cheng, K.K.Y., BiglarBaygi, A., & Solaymani, M. (2009). The effect of text authenticity on the performance of Iranian EFL students in a C-test. *Research in Language*, 7, 61-74.
- Chehrzad, M.H., & Ajideh P. (2013). Effects of different response types on Iranian EFL test-takers' performance. *Iranian Journal of Applied Language Studies*, 5 (2), 29-50.
- Combe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners*. U.S.A: The University of Michigan.
- Fountas, I., & G.S. Pinell. (2001). *Guiding readers and writers grades 3-6: Teaching comprehension, genre, and content literacy*. Portsmouth, NH: Heinemann.
- Grabe, B. (2002). Using discourse patterns to improve reading comprehension. *JALT*, Shizouka Conference proceedings. Tokyo: Japan Association for Language Teaching.
- Jiang, X. (2012). Effects of discourse structure graphic organizers on EFL reading

- Kendeou, P., Van Den Broek, P. (2007). The effect of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35 (7), 1567-1577.
- Kobayashi, M. (2002) Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, 19 (2), 193-220.
- Kobayashi, M. (2004). Investigation of test method effects: text organization and response formats: A response to Chen. *Language Testing*, 21 (2), 235-244.
- Liu, J. (1998). The effect of test methods on testing reading. *Foreign Language Teaching and Research*, 2, 48-52.
- Liu, F. (2009). The effect of three test methods on reading comprehension. An experiment. *Asian Social Sciences*, 5 (6), 147-153.
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399-438.
- Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In McNamara, D.S. (Ed.), *Reading comprehension strategies: Theories, interventions, Technologies*, (107-136). U.S.A. New York: Lawrence Erlbaum Association.
- Meyer, B. J. F. (1975). *The organization of prose and its effect on memory*. Amsterdam: North-Holland.
- Meyer, B. J. (1985). Prose analysis: Purposes, procedures, and problems. In Briton, B. K., Black, J.B. (Eds.), *Understanding expository text*, (269-297). Hillsdale, NJ: Erlbaum.
- Meyer, B., & Poon, L. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology*, 93, 141-159.
- Paltridge, B. (1996). Genre, text type, and the language learning classroom. *ELT Journal*, 50 (3).
- Rahimi, M. (2007). L2 reading comprehension test in the Persian context: Language of presentation as a test method facet. *The Reading Matrix*, 7 (1), 151-165.
- Rauch, D. P., & Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, 52 (4), 354-379.
- Rupp, A.A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23 (4), 441-474.
- Sadighi, F., Yamini, M., & Ayatollahi, M.A. (2007). Effects of text structure on reading. *Modern Language Journal*, 65, 43-53.
- Salmani Nodoushan, M.A. (2010). The impact of formal schemata on L3 reading validity of the tests in Iranian EFL students. *Higher Education of Social Science*, 2 (1), 16-20.

- Sharp, A. (2002). Chinese L1 schoolchildren reading in English: The effects of rhetorical patterns. *Reading in a Foreign Language*, 14 (2), 111-135.
- Sharp, A. (2004). Strategies and predilections in reading expository texts: The importance of text patterns. *RELC Journal*, 35 (3), 329-349.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147-170.
- Snyder, A.E. (2012). *The effects of graphic organizers and content familiarity on second graders' comprehension of cause and effect* (Unpublished doctoral dissertation): Columbia University, U.S.A.
- Sun, X. (2001). Do the different test methods have an effect on reading comprehension scores? *Primary and Middle School English Teaching and Research*, 3, 26-28.
- Tavakoli, M., Ahmadi, A., & Bahrani, M. (2011). Cloze test and C-test revisited: The effect of genre familiarity on second language reading test performance. *Iranian Journal of Applied Linguistics (IJAL)*, 14 (2), 173-204.
- Williams, H.P., Pollini, S., Nubla-kung, A.M., Snyder, A.E., Garcia, A., Ordynans, J.G., & Atkins, J.G. (2014). An intervention to improve comprehension of cause/effect through expository text structure instruction. *Journal of Educational Psychology*, 106 (1), 1-17.
- Wolf, D.F. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern language Journal*, 77 (4), 473-489.
- Yoshida, M. (2012). The interplay of processing task, text type, and proficiency in L2 reading. *Reading in a Foreign Language*, 24 (1), 1-29.
- Zhang, X. (2008). The effect of formal schema on reading comprehension: An experiment with Chinese EFL readers. *Computational Linguistics and Chinese Language Processing*, 13 (2), 197-214.