



The Effect of Content Familiarity & Test Format on Iranian EFL Test Takers' Performance on Test of Reading Comprehension

Omid Rezaei (Corresponding author)

Faculty of Foreign Languages, University of Isfahan, Isfahan 8174673441, Iran

E-mail: omidrezaei.rezaei99@gmail.com

Hossein Barati

Faculty of Foreign Languages, University of Isfahan, Isfahan 8174673441, Iran

Manijeh Youhanaee

Faculty of Foreign Languages, University of Isfahan, Isfahan 8174673441, Iran

Received: 11-06- 2012

Accepted: 30-06- 2012

Published: 01-09- 2012

doi:10.7575/ijalel.v.1n.4p.1

URL: <http://dx.doi.org/10.7575/ijalel.v.1n.4p.1>

Abstract

This study investigated the effect of content familiarity and test format on Iranian English learners. The participants of this were advanced students studying at different language institutes in Isfahan, Iran. To sample the subjects of this study, the latest version of Oxford Placement Test was administered to 428 students studying at advanced level in 6 different language institutes. Based on the results of the OPT test and for the sake of homogeneity 70 students were considered as the target participants of the study. Each participant was given a test of reading comprehension with familiar content and unfamiliar content. Each test contained multiple choice, true/false, and fill in the blanks test items. Factorial design results indicated that test takers had a significantly better performance on content familiar tests and sub tests. It also became clear that their performance on multiple choice section either in content familiar and content unfamiliar test was superior to that of true/false and fill in the blanks. It will be of endless help to test makers and language teachers to be aware of the role test format and content of the test can play on test takers' performance.

Key words: reading comprehension, schema theory, content familiarity, test format

1. Introduction

Reading is the most important academic skill (Carrell 1988a; Grab & Stoller 1996). The reading ability has always been viewed as critical to academic success (Bernhardt, 1991). Research on reading has attempted to look for components that affect reading performance as well as reading behaviors that distinguish proficient from less-proficient readers.

Text comprehension is a complex cognitive skill in which the reader should construct meaning by using all the available resources from both the text and background knowledge. These resources assist readers in utilizing lexis and syntax, retrieving their meaning from one's mental lexicon, making inferences and employing schemata (Alderson, 2000). Correct implementation of psychological recourses can help readers in the successful comprehension of the text (Yazdanpanah, 2007).

One of these psychological resources is background knowledge. Background knowledge has been discussed as a part of content schemata. Schema is the technical term used by cognitive scientists to describe how people process, organize, and store information in their heads. Schemas, or schemata, are seen as cognitive constructs by which we organize information in our long-term memory (Widdowson, 1983). Schemata, therefore, have been called "the building blocks of cognition" (Rumelhart, 1982) because they represent elaborate networks of information that people use to make sense of new stimuli, events, and situations.

According to Brown (2001) the hallmark of schema theory, with regard to reading, is that a text does not by itself carry meaning. The reader brings information, knowledge, emotion, and culture – that is schemata, to the printed



word. Clark and Silberstein (1977, quoted in Brown 2001) indicate that research has shown that reading is only incidentally visual. More information is contributed by the reader than by the print on the page. Schema theory gives perception into how the text and the reader's previous knowledge interact by often activating information that is relevant to the problem to be solved (Rumelhart, 1988; Nassaji, 2002). According to schema theory, our system contains an enormous number of schemata including all the daily activities and our experiences and observations (Rumelhart & Ortany, 1977 p: 128).

This would all seem to point to the fact that our understanding of a text depends on how much related schema we, as readers, possess while reading. Consequently, readers' failure or confusion to make sense of a text could be caused by their lack of appropriate schemata that can easily fit with the content of the text. This lack of appropriate schemata can be either formal or content-based. Brown (2001) defines these two as follows: content schemata includes what we know about people, the world, culture, and the universe, while formal schemata consists of our knowledge about discourse structure. The very important role of background knowledge on reading comprehension is noted by Carrell and Eisterhold (1983) and Anderson (1999) that a reader's comprehension depends on her ability to relate the information that she gets from the text with her pre-existing background knowledge.

Enquiries into the schema-comprehension relationship have been conducted initially and primarily in the realm on English as a first language. In general, studies in first language reading have pointed out the critical role of knowledge on comprehension (Koda, 2005; Recht & Leslie, 1988; Johnston, 1984). The application of schema theory to second language reading dates back to the 1980's (Carrell, 1981, 1983a, 1983b; Hudson, 1982; James, 1987, and Johnson, 1981) with studies conducted with ESL students of many different instructional levels.

In a paper by Swaffer (1988), it is mentioned that schema knowledge can be more influential in reading comprehension than word knowledge. Further in the article, she claims that content familiarity facilitate language recognition, recall of concepts and inferential reasoning (p: 126). Afflerbach (1986) showed that familiarity with the reading content enhance the reconstruction of the main idea. Brantmeier (2003) emphasizes that content familiarity can be an extremely significant factor in affecting L2 comprehension.

In addition to content familiarity the test itself may affect the readers' performance. Reading ability is examined in a form of reading comprehension test. Reading comprehension tests involve such skills as skimming vocabulary, recognition, guessing vocabulary meaning from the context cues, understanding the syntactic structure of the passage, grasping the main idea of the passage (Anderson, 1990). Reader's level of mastery in text comprehension can be assessed by different measures. Some reading test forms are multiple choice, open-ended questions, cloze tests, true/false, fill in the blank, written recall, sentence completion, matching activities and checklist tests (Farhady, 1987). Based on the ability measured in each question, the reader adopts a particular approach such as top-down, bottom-up or both. In other words different test items demand different strategies on the part of the reader (Yazdanpanah, 2007).

Accordingly, Anderson et al (1991) found that test items affect examinees' responses and their interaction with the text. Some items required test takers to reread parts of the passage, process deep level comprehension or scan, while some items merely required a surface level understanding of the passage.

2. Background of the study

The role prior knowledge plays in language comprehension has been formalized as the schema theory (Bartlett, 1932; Rumelhart, 1980). The term schema, as it is currently used, was introduced by Frederic Bartlett in his book, *Remembering*, in 1932, in which he described schemata as general impressions. The study of the influence of prior knowledge structure or schemata on comprehension and recall is called the schema theoretical research (James, 1988). Based on what James (1988) described schema theory, the process of comprehension is under the idea that input is put upon preexisting knowledge so that a match is made.

The thinking of Gestalt psychologists studying mental organization in the 1930s shows the stressing of whole rather than individual parts. This suggests that if some individual part(s) is/are missed for any reason, the organization of the whole can compensate for it. All the preexisting knowledge stored in one's mind can act as a whole. If, while reading, some element (s) in the reading passage is/are unknown, by anchoring the reading passage to what has been previously acquired in mind, the misunderstanding will be removed.

Later in the century, David P. Ausubel (1983) promoted the idea of the importance of prior knowledge in learning by stating that new propositions in the texts are anchored to general concepts that are stored in the mind (Anderson and Pearson, 1984).



Based on Rumelhart (1981), Schema is defined as a knowledge structure that shows aspects of our total knowledge which has been stored in our long term memory. It is defined as an abstract knowledge structure that shows the relationships among its components (Anderson and Pearson , 1984) , and an organization in the mind of an individual regarding the past experiences s/he has had (Barlett , 1932) . It is a prototype of some mental concepts that may represent various kinds of object, animal, people, abstract ideas, actions, or events and encompasses a network of knowledge related to a particular concept which may comprise additional schemata (Rumelhart , 1980). Schemata are defined by Widdowson (1983) as cognitive capacities which allow the information to be organized in long-term memory and provide a basis for production and as knowledge structures which are highly complex (Va Dijk, 1980). They, therefore, serve as devices for categorizing and arranging information so that it can be interpreted and retained.

Learning proceeds primarily from prior knowledge, as it is shown by a large body of finding, and only secondarily from the presented written materials .For example, the information used for filling in gaps in a cloze test passage and expecting further inputs does not come from the individual sentences but from the overall bodies of knowledge to which they refer .The act of picking out a pair of shoes is getting dressed in daytime as opposed to a part of buying the apparel. Therefore, memory must be organized in a way that the relevant general or specific information can be accessed from the input materials because prior knowledge influences learning and the learners construct concepts from prior knowledge.

A growing body of empirical research has investigated the influence of content familiarity on reading comprehension (Abu-Rabia, 1996, 2003; Carrell, 1987; Droop & Verhoeven, 1998; Johnson, 1981; Pritchard, 1990; Steffensen & Joag-dev, 1992; Steffensen, Joag-dev, & Anderson, 1979, Yeut Hung Chan, 2003, among others). In terms of methodology, most studies investigating the issue have included at least two groups of subjects. One group is provided with a content familiar text, the other one with a content unfamiliar text. Results of these studies indicate that readers who read content familiar text outperformed those that read unfamiliar text.

Hyde and Linn (1988) contended that the lower scores of women in the United states on the language part of the American Scholastic Aptitude Test (SAT) were mainly attributed to changes in the content of the reading of the test. This claim was in line with research that supported the assertion that in college-level achievement tests successful reading is related to the passage content from which exam questions are developed. In a study on gender differences in achievement test performance at the college level, Doolittle and Welch (1989) found notable gender differences for items associated with specific passage reporting that females scored higher than males with humanities-oriented reading passages, but lower than males with science-oriented passages.

The native language reading research with readers from different groups with the same national identity paralleled a similar study in L2 reading. Young and oxford (1997) investigated the differences between English speaking men and women while reading two Spanish texts and one English text. The texts included content such as, economics, the presence of foreign culture in work, leisure and history with regard to the recall scores, no significant differences by gender were reported for all three texts.

In a similar study, Brown (1982) administered an engineering reading test consisting of three reading passages to 116 college students at UCLA. Results showed that engineering students performed better than nonengineering students on items involving both specific engineering knowledge and general engineering content. Erickson and Molloy (1983) conducted a similar study based on a reading test that was also administered to a group of 83 college students. They were able to confirm Brown's finding that engineers significantly outperformed nonengineers with regard to engineering content, in both specific and general engineering reading. Similar findings were also reported by Alderson and Urquhart (1984) that engineering students ($n = 11$) performed better on engineering-related reading passages and economics students ($n = 11$) performed better on economics-related passages.

The sample sizes of the above studies are relatively small or modest. Hale (1988) examined the impact of major-field area on reading performance with a larger sample. He examined data from 32,467 graduate school applicants from four TOEFL paper-and-pencil administrations and found that students in two key major-field groups, the humanities/social sciences and the biological/physical sciences, performed significantly better on passages involving content relevant to their majors than on other passages. Although differences were statistically significant, the practical effect sizes were found to be small. The sources of the text could account for the small effect sizes. Hale hypothesized that because the texts were drawn from general readings, the advantage of studying a particular major was not as great as it could have been had the texts been drawn from specialized textbooks.



One specific subclass of knowledge of the world (i.e., content schemata) is "cultural knowledge". As Alderson (2000) states the knowledge of the world is idiosyncratic. It refers to every specific person's world. Our world is different from our friends' because every person has unique personal history, feelings, ideas, interests, and experiences not necessarily experienced or possessed by others. Cultural influences can also affect test takers' reading performance. Keshavarz, Atai, and Ahmadi (2007) investigated the contribution of content and background knowledge, vocabulary and syntactic knowledge, and L2 proficiency to reading comprehension and recall. The participants were 240 male Iranian students who learned English as a foreign language. Each participant was tested with two types of texts: an extract from the biography of an Islamic religious leader who is supposed to be familiar to the Muslim participants, and an extract from the biography of a non-Islamic religious figure. The authors found that familiarity with content was significantly correlated with reading comprehension test scores and recall scores ($p < .000$).

Similarly Chihara, Sakurai, and Oller (1995) based on the assumption that "very simple things like nouns referring to persons and places carry with them some fairly subtle semantic and pragmatic information" (p. 144), they changed several culturally unfamiliar words (e.g. Nicholas, Athen, Klein) from two English texts into more familiar words (Ben, Osaka, Daiei) for the Japanese participants. Leaving all other words intact, Chihara et al. could control other possible intervening variables such as the content and syntactic complexity of the two texts. The results of the cloze test showed that the participants performed significantly better on the modified, that is culturally familiar, cloze texts than on the original texts.

Chihara et al.'s (1989) study was later followed by Al-Fallay (1994, as cited in Oller, 1995), who examined the effect of cultural familiarity on Arab EFL students' cloze test performance over a three month period. In his study, he used two narratives. One was an English translation of a story originally written about an Arab setting, and the other was a story in English by an American with reference to events that took place in the United States. The American story was adapted positively to fit Arab expectations, and the Arab story was adapted negatively to violate Arab expectations. Five cloze tests over the Arab story and five over the Arabized version of the American story were presented sequentially in ten administrations over a three month period to the experimental group. The counterpart tests, consisting of five tests over the unmodified American story and five tests over the Americanized version of the Arab story were presented over the same period in matched administrations to the control group. Results indicated that the experimental group performed better on all ten cloze tests owing to the conformity of the textual material to their sociocultural expectations.

Sasaki (2000) investigated how schemata activated by culturally familiar words might have influenced students' cloze test-taking processes. Two groups of Japanese EFL learners with equivalent English reading proficiency completed either a culturally familiar or an unfamiliar version of a cloze test. Partially replicating Chihara et al.'s (1989) experiment, she changed several unfamiliar words in the original cloze test passage to more familiar ones in the modified version. Unlike Chihara et al., however, students were asked to give verbal reports of their test-taking processes, and to recall the passage after they had completed the test. Results demonstrated that those who read the culturally familiar cloze text tried to solve more items and generally understood the text better, which resulted in better performance, than those students who read the original text.

In another study, Floyd and Carrell (1987) designed an experimental study for 34 intermediate-level ESL students attending a college-level English program. Participants in the treatment group received two training sessions on cultural background knowledge. Pre- and post- culture-related reading tests were used to measure any potential change in reading ability for the treatment and control groups. The authors reported that students in the treatment group performed significantly better than those in the control group on passages containing pertinent cultural information.

Brantmeier (2002) employed authentic narratives passages and comprehension assessment tasks to two groups of students from advanced university grammar and literature courses. Across both levels, male participants reported being more familiar with the topic of boxing than the females did, and female participants indicated being more familiar with the topic of a frustrated housewife than the males did. Furthermore, results showed no significant gender differences in comprehension of the gender-oriented passages at the more advanced stages of acquisition.

Brantmeier (2003) reported significant interactions between readers' background knowledge and comprehension among intermediate second language learners of Spanish at the university level. The two passages employed in this study were authentic narratives, which were employed in the previous research, about a boxing match and a frustrated housewife. It also indicated a significant gender difference with comprehension assessed via multiple choice items as well as written recall. Males scored higher on both test forms with the boxing match text content.



This study provided evidence that readers' gender and passage content interact in ways that affect second language reading comprehension.

In addition research in language testing has also pointed out that test-takers with different characteristics might be affected by a test in ways that are not relevant to the abilities being tested (Bachman, 1990; Kunnan, 1998). Test format has been shown to be an important facet that could influence different test-takers' test performance (Bachman & Palmer, 1982; Shohamy, 1984, 1997). The issue of test format differences has been the subject of debate because it is generally assumed that different test formats elicit different levels of skills or abilities; therefore, such tests are subject to having different effects on test-takers from various linguistic and cultural backgrounds. Kunnan (2004) raised the issue of test fairness, arguing that certain test formats may favor some groups of test-takers but not others, threatening the validity of a particular test. Shohamy (1997) claimed that language tests employing test methods that are unfair to different groups of test-takers are unethical. If group performance differences do exist, the reason should be real differences in the skills or abilities being tested instead of confounding variables such as test format (Elder, 1997).

A variety of test formats or methods have been employed in language testing, including cloze, c-test, gap-filling, matching, multiple-choice, open-ended (or short-answer), ordering, recall, summary, and summary gap-filling (e.g., Alderson, 2000; Buck, 2001; Kobayashi, 2002). Since there is no perfect test format that functions well in every situation, researchers must understand the characteristics of each format and make the best selection according to which one(s) most appropriately serve(s) the purpose of a test in each context. While the literature on test format effects is enormous, the literature review below focuses on comparing multiple-choice and open-ended formats e.g. fill in the blanks in reading and listening.

This has been one of the most investigated comparisons; thus, it provides a solid basis for quantitative synthesis. Based on Davies et al. (1999), these two formats are defined as follows. Multiple-choice is a format with a stem and three or more options from which learners are required to select one. An open-ended format refers to a question that requires learners to formulate their own answers with several words or phrases.

The previous literature from a quantitative perspective has mainly focused on two issues: (a) differences in construct or trait measured using multiple-choice and open-ended formats and (b) differences between test scores in multiple-choice and open-ended formats (i.e., the relative difficulty of test formats). Among a wide range of studies investigating the former issue (e.g., Bennett & Ward, 1993; Buck, 2001; Campbell, 1999; Cohen, 1998), highly important and interesting in terms of its comprehensive coverage of previous studies is Rodriguez (2003), which synthesized 56 sets of correlation coefficients between stem-equivalent multiple-choice and open ended formats based on 29 studies from a variety of disciplines. The results revealed a correlation between them approaching unity (0.95 [95% confidence interval: 0.91, 0.97]). This appears to suggest that multiple-choice and open-ended formats measure a very similar construct when they use the same stem. In the case of the differences between the test scores in both test formats, most studies have used the t test or analysis of variance to compare the mean scores in L1 reading, L2 reading, and L2 listening. Since format effects vary by domain (Traub, 1993), the previous findings are summarized below.

The literature in L1 reading has provided mixed results. Some studies have shown that multiple-choice formats are easier than open-ended formats (e.g., Arthur et al., 2002; Davey, 1987), whereas some other studies have found no statistical difference between the two formats (e.g., Elinor, 1997; Pressley et al., 1990). In L2 reading, most studies have shown that multiple-choice formats are easier than open-ended formats (e.g., Shohamy, 1984; Wolf, 1991), which contradicts Elinor (1997) and Trujillo (2006), in which the two formats were considered to be of similar difficulty. All studies have shown that in L2 listening, multiple-choice formats are easier than open-ended formats (e.g., Berne, 1992; In'nami, 2006; Teng, 1999). What became clear through the review of L1 reading, L2 reading, and L2 listening was that in most studies, multiple-choice formats were easier than open-ended formats; however, in some L1 and L2 reading studies, no significant difference was found between the two formats.

2.1 Statement of the problem

Having proper information about the relationship between EFL learners' prior knowledge and their reading comprehension will help finding a useful way to identify the appropriate text readability for the students at different levels. The study of the impacts of EFL learners' prior knowledge on comprehending texts has usually been conducted through experiments about L2 learners' reading performance. Little has been reported in the related literature about EFL learners' perceptions of their prior knowledge in terms of its contributions to reading comprehension.



In addition to the effect prior knowledge has on learners' performance on reading comprehension tests, test format itself also proved to be a significant factor influencing their performance. The research literature reports that confounding factors in testing reading comprehension such as varying test formats may differentially affect the performance of students from diverse backgrounds. This project is an effort made to find out how test format along with content familiarity affects Iranian EFL learners' performance on reading comprehension tests of various form.

2.2 Research question

The present study intended to address the following questions:

1. Does content familiarity have a significant effect on EFL learners' performance on tests of reading comprehension?
2. Does test form have a significant effect on EFL learners' performance on tests of reading comprehension?
3. Does the interaction of content familiarity and test form have a significant effect on EFL learners' performance on tests of reading comprehension?

Based on the above mentioned questions following hypotheses were formed:

1. Content familiarity has no significant effect on Iranian EFL learners' performance on tests of reading comprehension.
2. Test form has no significant effect on Iranian EFL learners' performance on tests of reading comprehension.
3. The interaction of content familiarity and test form has no significant effect on EFL learners' performance on tests of reading comprehension.

3. Method

3.1. Participants

The population from which this study aimed to draw its sample was advanced students studying at different language institutes in Isfahan, Iran. To sample the subjects of this study, the latest version of Oxford Placement Test test was administered to 428 students studying at advanced level in 6 different language institutes. Based on the results of the OPT test and for the sake of homogeneity 70 students were considered as the target participants of the study. Based on the rating scale of the OPT test those who scored between 50 and 60 are considered as advanced learners. All the participants of the present study scored 56, so it seems homogeneity is highly observed.

The sample included both boys and girls, but due to the fact that the number of female learners is bigger, 65% of the participants were girls. For being respectful to research ethic a book was gifted to each participant of the study for his/her effort during the course of data collection.

3.2. Instruments

Two different types of texts were used in this study. One type was with familiar content, and the other with unfamiliar content. To minimize the effect of extraneous factors all the texts were of similar topics. This was provided by choosing texts about Iranian scientists as content familiar and texts about foreign scientists as content unfamiliar. The next section reports in more details what the characteristics of each group of texts were.

3.2.1. Content familiar tests

Texts with familiar content were about Iranian scientist namely Khaje nasiro din tousi, Professor Hessaby, and Abou reihane biruni. Texts were adopted from the internet and were analyzed for their authenticity. Some parts of the texts were removed to keep them with similar length; however care was taken not to damage the unity of the texts. The texts were then put into a test of reading comprehension. Each text was followed with various formats: multiple choice, fill in the blanks, and true/false. As a result, each text was followed by 12 items. Thus the total test had three texts followed by 36 items. This made the content familiar test for the present study.

3.2.2. Content unfamiliar tests

Content unfamiliar texts were about foreign scientists namely Feleming, Dalton, and Hawkins. Texts were adopted from the internet; this would presumably help the authenticity of the texts. The same procedure as discussed for the content familiar test was undertaken to make the length of the texts approximately similar and at the same time does not damage their unity and coherence. As discussed above each text was put into a test of reading comprehension. The same steps discussed for content familiar test were taken for content unfamiliar tests to make both tests similar in format but different in content. This made the content unfamiliar test for the present study

3.3. Procedure

Before embarking on the process of data collection for the main study, it was necessary to put the research instruments into piloting to make sure they function the way the researcher expected.

3.3.1. Pilot study

Both content familiar and content unfamiliar tests were piloted with 15 participants similar to those of the target group. The piloting phase of the study had 4 purposes: First determining the difficulty level of the texts, second determining the timing of the test, third determining the reliability estimates of the tests, and finally the clarity of the directions use in the test.

For estimating difficulty level of both content familiar and unfamiliar texts Fry index of readability was used. This indicated that all six texts were of the same level of difficulty, and according to Fry's graph belonged to the same grade level. Table 1 indicates the grade level of all six texts.

Table 1. Readability index of content familiar and content unfamiliar texts

Texts	syllables per 100 words	sentences per 100 words	Grade Level
Content familiar			
Khaje	158	5.3	9
Hessaby	167	4.9	9
Abo Reihan	159	5.1	9
Content unfamiliar			
Feleming	162	4.8	9
Dalton	154	4.4	9
Hawkins	166	5.4	9

According to Fry's readability graph, texts of the same grade level are of equal difficulty level. As it is clear from table 1 all six texts belong to grade level 9, so they are of the same difficulty level.

Each text was reviewed by five judges with experience of teaching English for many years to check the level of difficulty and appropriateness of the texts. During the phase of piloting students showed no sign of confusion, so it indicates that tests were within their level of proficiency.

Table 2 indicates the timing of both content familiar and unfamiliar tests.

Table 2. Timing for content familiar and content unfamiliar tests

text	timing
Content familiar	
Khaje	15 m
Hessaby	15 m
Abo Reihan	18 m
Content unfamiliar	
Feleming	20 m
Dalton	20 m
Hawkins	18 m

For checking the internal consistency of Cronbach's alpha was run. Table 3 indicates the reliability estimates for both content familiar and content unfamiliar test.

Table 3. Cronbach's alpha results for content familiar and content unfamiliar tests

Texts	True/False	Fill in the blanks	Multiple Choice	Total
Content familiar	0.68	0.74	0.88	0.73
Tousi	0.76	0.81	0.78	0.77
Biruni	0.76	0.85	0.76	0.80
Hessaby	0.80	0.77	0.78	0.76
Content unfamiliar	0.71	0.78	0.83	0.71
Dalton	0.79	0.76	0.77	0.75
Fleming	0.76	0.77	0.83	0.78
Hawkins	0.86	0.74	0.78	0.80

According to George & Mallery (2003), acceptable value of Cronbach's alpha is 0.64 and higher. Thus as table 3.3 indicates that all the tests along with their sub sections enjoyed acceptable levels of reliability.

Almost all test takers expressed no confusion about the directions used in the tests. This means that tests directions were quite clear for them; however some of the directions were simplified since they were misunderstood by some students.

3.3.2. Main study

For the main study participants were asked to take three content familiar and unfamiliar tests in two different sessions which were held in the morning and afternoon respectively. They were fully informed of the importance of their contributions to the results of the study and were asked to pay due attention and discretion and patience while taking the tests. In order to decrease the chance of cheating the tests items were arranged in three different orders. There was a five hour time interval between the two test sessions. The test results were emailed to the students. To correct the fill in the banks section questions, the "acceptable word correction method" was used.

4. Result

In the following the results of the statistical procedure utilized for answering each research question will be discussed.

4.1. The effect of content familiarity on test takers' performance

The first research question addressed in this study was:

1. Does content familiarity have a significant effect on Iranian EFL learners' performance on tests of reading comprehension?

To address this question the mean of the scores on the content familiar section, consisting of three test, and content unfamiliar section, consisting of three tests were obtained. Table 10 indicates the descriptive analysis of the above mentioned scores.

Table4. Descriptive analysis of scores of content familiar and content unfamiliar tests

	Content Familiar	Content Unfamiliar
Mean	9.8048	7.7286
Std. Deviation	.55166	.76201
Variance	.304	.304
Skewness	.076	.076
Kurtosis	.056	.056

As indicated in Table 4.1, the mean scores for content familiar and content unfamiliar tests were 9.8048 and 7.7286, respectively. The mean scores indicated that students' performance on content familiar tests was superior to their performance on content unfamiliar tests.

The data collected for addressing the first research question of this study were then put into inferential analysis. Table 4.2 presents the results of T-test.

Table 5. T-test results for content familiar and content unfamiliar tests

	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
	Lower	Upper			
multiple choice familiar - multiple choice unfamiliar	-2.29487	-1.85751	-18.940	69	.000

The results of T-test in table 4.2 indicate that there was a statistically significant difference between test takers' performance on content familiar and content unfamiliar tests. According to the t value the null hypothesis is rejected.

4.2 The effect of test form on test takers' performance on reading comprehension tests

The second question addressed in this study was:

2. Does test form have a significant effect on EFL learners' performance on tests of reading comprehension?

To address this question a factorial designs was run on the scores obtained from multiple choice, true/false, and fill in the blanks sections of both the content familiar and content unfamiliar tests. For each test form the mean score of all 6 tests were estimated then was compared to that of two other forms. By doing this, the effect of content familiarity was kept minimized and all tests were considered equal. Table 4.9 below indicates the results of the factorial design.

Table 6. Factorial design results on scores of multiple choice, true/false, and fill in the blanks sections of all tests

(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig. ^a
multiple	True/False	.376*	.047	.000
	Fill in the blank	.557*	.047	.000
True/False	multiple	-.376*	.047	.000
	Fill in the blank	.181*	.047	.000
Fill in the blank	multiple	-.557*	.047	.000
	True/False	-.181*	.047	.000

As it can be seen in Table 4.9, test takers' performance on multiple choice items was significantly different from their performance on true/false and fill in the blanks items.

4.3 The effect of the interaction of content familiarity and test form on test takers' performance on reading comprehension tests

The third question addressed in this study was:

3. Does the interaction of content familiarity and test form have a significant effect on EFL learners' performance on tests of reading comprehension?

To address this question a factorial design was run on the scores obtained from test takers' performance on multiple choice, true/false, and fill in the blank sections of tests with familiar and unfamiliar content. Table 4.10 indicates the results of factorial design of the above mentioned scores.

Table 7. Factorial design results on scores of multiple choice, True/false, and fill in the blank section of tests with familiar and unfamiliar content

Test	(I) Type	(J) Type	Mean Difference		
			(I-J)	Std. Error	Sig. ^a
familiar	multiple	True/False	.276*	.067	.000
		Fill in the blank	.533*	.067	.000
	True/False	multiple	-.276*	.067	.000
		Fill in the blank	.257*	.067	.000
	Fill in the blank	multiple	-.533*	.067	.000
		True/False	-.257*	.067	.000
unfamiliar	multiple	True/False	.476*	.067	.000
		Fill in the blank	.581*	.067	.000
	True/False	multiple	-.476*	.067	.000
		Fill in the blank	.105	.067	.119
	Fill in the blank	multiple	-.581*	.067	.000
		True/False	-.105	.067	.119

As it is shown in Table 4.9, test takers' performance on the multiple choice, true/false, and fill in the blank section of tests with familiar content was significantly different at .000 level of significance. In each time test takers' performance on one form e.g. multiple choice, is compared with two other forms e.g. true/false and fill in the blank. In all comparisons test takers' performances were significantly different. But in case of tests with unfamiliar content test takers' performance on multiple choice section was significantly different from true/false and fill in the blanks section. On the contrary, their performance on true false and fill in the blank sections was not significantly different.

Figure 4.1 will illustrate test takers' performance on content familiar tests and their sub sections in comparison to content unfamiliar tests and their sub sections.

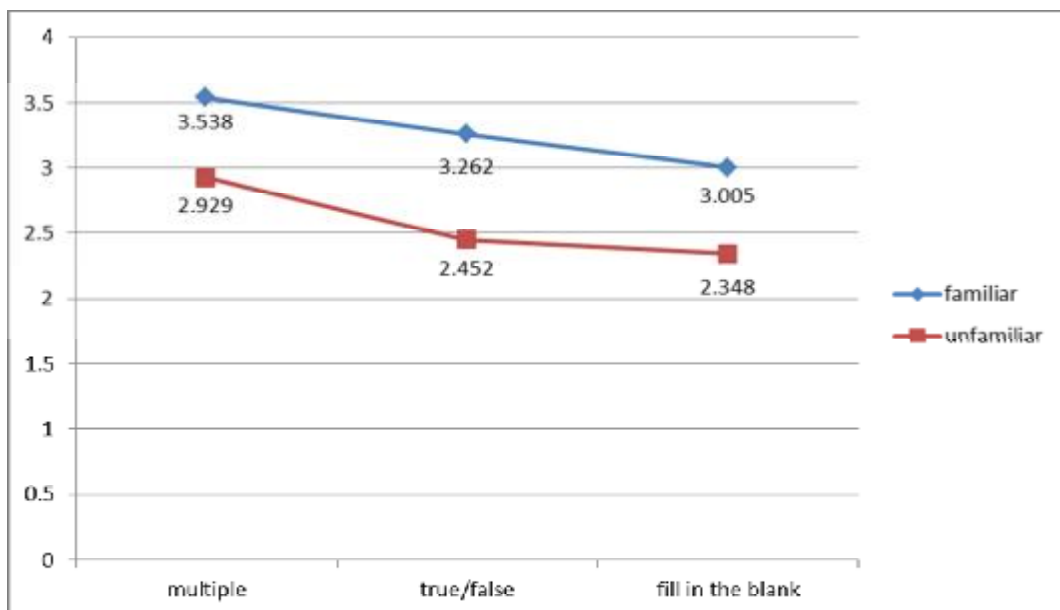


Figure 1. Estimated marginal means of scores



As it can be seen in Figure 4.1, test takers' performance on content familiar tests was superior to their performance on content unfamiliar tests. Regarding different test forms also it can be inferred that test takers' performance on different sub tests of content familiar tests was superior to those of content unfamiliar tests.

It can also be observed that test takers' performance on multiple choice section of both content familiar and content unfamiliar tests was significantly better than that of true/false and fill in the blank sections. In content familiar tests, their performance on true/false section was significantly better than that of fill in the blank section. On the other hand in content unfamiliar tests their performance on true/false section was not significantly better than that of fill in the blank section.

5. Discussion

This study aimed at investigating the effect of content familiarity and test format on Iranian EFL test takers' performance on reading comprehension tests. The results confirmed what has been widely acknowledged as the positive effect of background knowledge and content familiarity on reading comprehension (e.g., Alderson, 2000; Alptekin, 2006; Ketchum, 2006; Oller, 1995; Pulido, 2003; Steffensen et al., 1979). The difference between the participants performances in comprehension suggested a strong possibility that when students read the content familiar texts possessed relevant background knowledge, which reduced the cognitive load imposed by the complex reading procedures (Perfetti, 1985) on the memory system (Baddeley, 1997; Ellis, 2001; Kintsch, 1998; McLaughlin et al., 1983), as opposed to the time when students had to deal with unfamiliar content and visualize the script in their minds. In the content familiar version, where the text were about Iranian scientists, the students seemed to find it easier to allocate attentional resources to more linguistic elements and construct mental representations of the familiar context. In contrast, it was much more difficult for the students to create mental representations of the unfamiliar context, which prevented the economical use of attentional resources.

The effect of content familiarity may also be related to motivational issues, although studies in this area are few. Recent work on attitudes and motivation has shown that motivation is related to achievement (e.g., Dörnyei, 2003), involving many factors such as ownership and interest (Williams & Burden, 1997). In the realm of reading research, two types of interest have been proposed as contributing to motivation, namely, personal interest and situational interest (Alexander & Jetton, 2000). The first is related to one's general approach to reading while the latter is generated by the text. The latter category can be of practical value in explaining the variable performances of the students in this study.

When students read content familiar text, they might have enjoyed the situational interest aroused by the text. Tomlinson (1998), for example, maintained that when students see elements of their local culture in classroom materials, they feel much more engaged and identify themselves with the context of the text. Such a personal appeal has also been shown as one of the sources of situational interest (Schraw, Bruning, & Svoboda, 1995). Thus, students reading content familiar texts might have found the version much more appealing than content unfamiliar texts, which had no local reference at all. Such an appeal might, in turn, have contributed to higher levels of motivation (Dörnyei, 2003, 2005; Harter, 1981; Williams & Burden, 1997) and thus led to a better reading comprehension (e.g., Guthrie, Wigfield, Metsala, & Cox, 1999; Wigfield, 1997).

Besides, statistical results indicates that a particular test format appear to be easier than others in this case multiple choice. This ease may be the result of more frequent exposure of test takers to multiple choice format in Iranian context. In most standardized tests and even non-standardized tests the most frequent format of question is multiple choice. So it can be the reason why participants performed significantly better on multiple choice section of both content familiar and content unfamiliar tests.

As it can be seen from the results in the content familiar section test takers' performance on true/false section was superior to that of fill in the blanks, but this difference is not significant in the content unfamiliar tests. It can be said that the effect of content familiarity may prove to be more profound on the true false section.

6. Implications and conclusion

Two conclusions, with some caveats described below, can be drawn from this study. Firstly, content familiarity facilitates comprehension. Secondly, although test format affects reading comprehension, the influence of cultural familiarity remains intact. Therefore, if readers lack the relevant schema, various test forms cannot fully compensate for the discrepancy or help readers comprehend a text.

This study is not without limitations. Although efforts were made to homogenize the groups on the basis of their OPT scores, more standardized tests of proficiency such as TOEFL or IELTS could have contributed to more homogeneous sampling of the groups. Further, recent research on reading processes and the variables involved



suggests motivational and attitudinal factors are related to reading comprehension. These elements were not taken into account in this study. A better-controlled measurement could shed further light on the phenomenon. Finally, the study was conducted with a small sample of students. A larger sample could tolerate individual variations better in statistical analysis.

Despite these limitations, the results of this study have important pedagogical implications. Given the fact that the scores were higher with the content familiar tests, materials that contain references to local elements or have personal relevance can produce a facilitative influence and can be useful in language classes. However, it may not always be possible, nor is it always desirable, to use texts only with direct local or personal references. Thus, one cannot underestimate the importance of being familiar with the content in the process of reading.

This study can also be helpful to test makers. They can decide on the format of the questions based on the teaching context and teachers' expectation from the test takers. If it is expected that test takers score higher on reading comprehension tests multiple choice items will be a better alternative than true/false and fill in the blanks.

References

- Abu-Rabia, S. (1996). Attitudes and cultural background and their relationship to reading comprehension in a second language: A comparison of three different social contexts. *International Journal of Applied Linguistics* 24(5), 711-721.
- Afflerbach, P. (1986). The influence of prior knowledge on expert readers' main idea construction processes. (ERIC Document Reproduction Service No. ED 284 193).
- Alderson, J. C. (2000). *Assessing reading*. NY: Cambridge University Press.
- Alexander, P. A., & Jetton, T. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. 3* (pp. 285-310). Mahwah, NJ: Earlbaum.
- Alptekin, C. (2006). Cultural familiarity in inferential and literal comprehension in L2 reading. *System*, 34, 494-508.
- Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data scores. *Language Testing*, 8, 41-66.
- Anderson, Neil J. (1990) *Exploring Second Language Reading: Issues and Strategies*. BOSTON: Heinle & Heinle Publishers.
- Bachman, L.F., & Palmer, A.S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16, 449-465.
- Baddeley, A. (1997). *Human memory*. Hillsdale, NJ: Lawrence Erlbaum
- Bartlett, F.C. (1932). *Remembering: A study in experimental and social Psychology*. Cambridge: Cambridge University Press.
- Bennett, R. E., & Ward, W. C. (Eds.). (1993). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*. Hillsdale, NJ: Erlbaum.
- Brantmeier, C. (2003). Does gender make a difference? Passage content and comprehension in second language reading. *Reading in a Foreign Language*, 15(1), 1-23.
- Brown, H.D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NY: Addison Wesley Longman, Inc.
- Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- Campbell, J. R. (1999). *Cognitive processes elicited by multiple-choice and constructed-response questions on an assessment of reading comprehension*. (UMI No. 9938651).
- Carrell, P. L. (1988). Content and formal schemata in ESL reading. *TESOL Quarterly*, 21, 462-481. Chihara, T., Sakurai, T., & Oller J. W. (1989). Background and culture as factors in EFL reading comprehension. *Language Testing In Oller, J.W. Jr. and Jonz, J., editors, Cloze and coherence*. London: Associated University Press, 135-47.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. Harlow, Essex, UK: Longman.



- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge, UK: Cambridge University Press.
- Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research and applications. *Language Learning*, 53, 3–32.
- Dörnyei, Z. (2005). *The psychology of the language learner*. Mahwah, NJ: Lawrence Earlbaum.
- Elder, C. (1997). What does test bias have to do with fairness? *Language Testing*, 14(3) 261-277.
- Elinor, S.-H. (1997, May). *Reading native and foreign language texts and tests: The case of Arabic and Hebrew native speakers reading L1 and English FL texts and tests*. Paper presented at the Language Testing Symposium, Ramat- Gan, Israel. (ERIC Document Reproduction Service No. ED 412746)
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge, England: Cambridge University press.
- Erickson, M., & Molloy, J. (1983). ESP test development for engineering students. In J. Oller, (Ed.), *Issues in language testing research* (pp. 280-288). Rowley, MA: Newbury House.
- Floyd, P. & Carrell, P. L. (1987). Effects on ESL reading of teaching content schemata. *Language Learning*.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. C. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, 3, 231–256.
- Hale, G. (1988). The interaction of student major-field group and text content in TOEFL reading comprehension (TOEFL Research Rep. No. RR-25). Princeton, NJ: ETS.
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientations in the classroom: motivational and informational components. *Developmental Psychology*, 17, 300–312.
- Hudson, T. (1982). The effects of induced schemata in L2 reading: Non-decoding factors in L2 reading performance. *Language Learning*, 32, 1-32.
- Hyde, J. & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104 (1), 53-69.
- In'nami, Y. (2006). *The effects of task types on listening test performance: A quantitative and qualitative study*. Unpublished doctoral dissertation, University of Tsukuba, Japan.
- James, A. (1988). *The acquisition of a second language phonology*. Tübingen: Gunter Narr.
- Keshavarz, M. H., Atai, M. R., & Ahmadi, H. (2007). Content schemata, linguistic simplification, and EFL readers' comprehension and recall. *Reading in a Foreign Language*, 19, 19-33.
- Ketchum, E. M. (2006). The cultural baggage of second language reading: An approach to understanding. *Foreign Language Annals*, 39, 22–42.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Erlbaum & Razi: The effects of cultural familiarity on reading comprehension
- Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, 19, 193–220
- Kunnan, A.J. (1998). Approaches to validation in language assessment. In A.J. Kunnan (Ed.), *Validation in language assessment* (pp. 1-16). Mahwah, NJ: Erlbaum.
- Kunnan, A.J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context: Proceeding of the ALTE Barcelona conference, July 2001* (pp. 27-50).
- McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning*, 33, 135–158. McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning*, 33, 135–158.
- Nassaji, H. (2002). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, 17, 85–114.



- Oller, J. W. (1995). Adding abstract to formal and content schema: Results of recent work in Peircean semiotics. *Applied Linguistics*, 16, 273–306.
- Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.
- Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, 53, 233–284.
- Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement*, 40, 163–184
- Rumelhart, D.E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, New Jersey: Erlbaum.
- Rumelhart, D.E. (1980). Schemata: The building blocks of cognition, Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, Lawrence Associates.
- Rumelhart, D.E. (1981). Schemata, the building blocks of cognition. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research review*. Newark, Delaware: I.R.A., 3–26.
- Rumelhart, E. E. (1982). Understanding. In J. Flood (Ed.), *Understanding Reading Comprehension*, (pp. 1–21).
- Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behaviour*, 27, 1–17.
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14(3), 340–349.
- Steffenson, M.S., Joag-dec, C., & Anderson, R.C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, 15, 10–29.
- Swaffer, J. K. (1988). Readers, texts, and second languages: the interactive process. *The Modern Language Journal*.
- Teng, H.-C. (1999, March). *The effects of question type and preview on EFL listening assessment*. Paper presented at the American Association for Applied Linguistics. (ERIC Document Reproduction Service No. ED 432920).
- Tomlinson, B. (1998). Introduction. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 1–24). Cambridge, England: Cambridge University Press.
- Traub, R. E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 29–44). Hillsdale, NJ: Erlbaum.
- Urquhart, A.H. (1984). The effect of rhetorical ordering on readability. In J.C. Alderson, & A.H. Urquhart (Eds.) *Reading in a foreign language*. London: Longman.
- Van Dijk, T.A. (1980). Story comprehension: An introduction. *Poetics*, 9, 1–21.
- Widdowson, H.G. (1983). *Language purpose and language use*. Oxford: Oxford University Press.
- Wigfield, A. (1997). Reading motivation: A domain-specific approach to motivation. *Educational Psychologist*, 32, 59–68.
- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers*. Cambridge, England: Cambridge University Press.
- Yazdanpanah, K.H. (2007). The effect of background knowledge and reading comprehension test items on male and female performance. *The Reading Matrix*. Vol. 7, No. 2
- Young, D. J. & Oxford, R. L. (1997). A gender based analysis used to process written output in the native and foreign language. *Applied Language Learning*, 8, 43–73.