THE EFFECT OF TEACHING WITHIN-TEXT KEY WORDS ON STUDENTS’ PERFORMANCE IN READING COMPREHENSION

Mohammad Reza Khodasenas
Applied Science and Technology, Iran
email: mrkhodashenas@yahoo.com

Abstract: The present study was conducted to investigate the effects of teaching within-text key word synonyms, opposites and related words on students’ performance on reading comprehension of TOEFL among Iranian EFL learners. To carry out the research, 60 Iranian EFL learners, who participated in a TOEFL preparation course, were selected as the participants of the study. Afterward they were randomly assigned into experimental and comparison groups. The experimental group was given a treatment including within-text key word synonyms, opposites and their translations, while the comparison group was given a placebo. To collect the required data, two instruments (a pre-test, and a post-test) were administered to both groups during the experimentation. Subsequently, students’ scores were collected through the administration of different tests and the results were statistically analyzed. The results of these analyses revealed that the experimental group outperformed the comparison group and thus, it was concluded that teaching within-text key word synonyms, opposites and related words can improve the reading comprehension ability and general proficiency of EFL language learners.

Key words: within-text key word synonyms, opposites, reading comprehension strategies.

PENGARUH PEMBELAJARAN KATA KUNCI DALAM TEKS ATAS KINERJA SISWA DALAM MEMBACA PEMAHAMAN

Abstrak: Kajian ini dilakukan untuk mengetahui pengaruh pengajaran katakunci sinonim, antonim, dan kata yang berkaitan dalam text terhadap kemampuan pemahaman bacaan TOEFL pembelajar EFL di Iran. Dalam kajian ini, 60 pembelajar EFL yang mengikuti kelas persiapan TOEFL dipilih sebagai peserta penelitian. Para peserta kemudian ditempatkan secara acak ke dalam kelompok eksperimen dan kelompok pembanding. Kelompok eksperimen diberikan perlakuan meliputi kata kunci sinonim, antonim, dan terjemahannya dalam text, sementara kelompok pembanding diberikan perlakuan berbeda (placebo). Untuk mendapatkan data kajian, dua instrument (pre-tes dan post-tes) diberikan kepada dua kelompok tersebut selama eksperimen. Kemudian, nilai-nilai para siswa dikumpulkan melalui pelaksanaan tes yang berbeda dan hasilnya dianalisis secara statistik. Hasil dari analisis ini membuktikan bahwa kelompok eksperimen lebih baik dari kelompok pembanding sehingga dapat disimpulkan bahwa pengajaran kata kunci sinonim, antonim, dan kata yang berkaitan dalam text dapat meningkatkan kemampuan pemahaman bacaan dan kecakapan umum pembelajar bahasa Inggris sebagai bahasa asing (EFL).

Kata kunci: kata kunci sinonim, antonim, strategi pemahaman bacaan, strategi mengikuti tes, validitas konstruk

Teaching English as a second or foreign language has become a very popular profession in today's world due to the fact that English has become a lingua franca of international communication. A great number and variety of English courses and
exams are available to those who desire to learn English. The teachers of English across the world use various techniques and methods to ensure that the students learn basic language skills such as listening, speaking, reading, and writing. Reading skill is the one which is sometimes neglected.

For many language learners reading is ranked first among the academic skills that they wish to gain mastery over (Grabe & Stoller, 2001; Richards & Renandya, 2002). This comes from the growing number of countries moving toward giving English instruction importance from a younger age (Ediger, 2001). It is believed that consolidating and extending one’s knowledge of language and gaining wide general information, world knowledge, as well as improving a skill like writing can take place by means of reading (Harmer, 2000; Rivers, 1981). Besides, it is reported that 10 to 15 percent of students experience difficulty in reading comprehension (Mercer, 2001). Reading is viewed to be understanding an unobservable internal and mental process (Ur, 1996; Vacca, Vacca, and Gove, 2000). The problem doubles when the skill is taken to be difficult to measure directly.

However, the issue of construct validity has been largely ignored because most current reading comprehension tests are breadth of knowledge measures which provide an estimate of how many words testees have in their lexicons. Spolsky (1985) suggests that language testing has evolved through three main phases: a first pre-scientific phase utilizing examinations subjectively marked by a single examiner, a second phase focusing on objectivity and reliability, and a third phase which emphasizes validity along with the aspects in the second phase. Unfortunately, in the case of reading comprehension tests, the evolution seems to have largely stopped at the second phase, and when the concept of validity has been addressed, it has typically not been construct validity, but other types such as criterion validity (Henning, 1991).

More recently, when the Test of English as a Foreign Language (TOEFL) was launched in 1964, vocabulary and reading comprehension items were combined in one section on the TOEFL. As a result of research by Henning (1991), the type of vocabulary item was changed in 1995 to one in which the target words were embedded in the reading comprehension passage with a four-option multiple-choice item attached. In fact this observation about the deficiency of construct validation can be made not just about TOEFL reading items, but about most other reading items and tests as well.

Test-Taking strategies for Reading
Reading is the most important of all skills for most language learners in general and for EFL learners in particular. Yet to date, no theory of reading has been proposed that is acceptable to all experts. As a result, tests of reading take various forms and evaluate different kinds of abilities such as pre-reading, reading speed, and reading comprehension. Pre-reading tests attempt to assess learners' ability to visually determine the printed characters of the language. Tests for reading speed gauge the speed with which the testees read quantities of material. And reading comprehension tests try to determine the examinees' ability to get meaning from printed material (Farhady, Ja'farpur, & Birjandi, 2000). According to Hughes (1989), receptive skills may also be difficult to assess because they are difficult to observe in overt behavior. According to Alderson (2000) it is important to understand that there is no one best method for assessment of reading, because assessment refers to a variety of ways of collecting information on a learner's language ability or achievement. Alderson believes that it is inadequate to measure the understanding of text by only one method, and objective methods can effectively be
supplemented by more subjectively evaluated techniques. He further holds that since readers typically respond to texts in a variety of different ways, it makes good sense to employ a number of different techniques even on the same text. Reading comprehension is often assessed through multiple-choice tests, true/false tests, cloze tests, and sentence completion tests. Reading comprehension tests are important instruments in psychological-pedagogical practice. Different versions of reading comprehension tests are part of admission or placement exams in a number of educational institutions. Reading comprehension tests have been frequently criticized regarding basic, fundamental test assumptions. Criticism mainly questions construct validity of these texts (Katz, Lautenschlager, Blackburn, & Harns, 1990).

The Nature of Instruction in Reading Comprehension
Reading is a complex developmental process which requires comprehension text to construct meaning from written text in relation to students' experimental and conceptual background, word recognition strategies and linguistic comprehension (Anderson, 1999). He believes that to enhance understanding of ideas, students will consciously apply strategies to particular constraints in the act of comprehending a particular text. For the readers to construct meaning they have to transact with text, by making meaning through the combination of prior knowledge and previous experience, information available in text, the stance they take in relation to the text, and immediate, remembered, or anticipated social interaction and communication. According to Pauline (2001) to promote comprehension and encourage reading as a lifetime habit, learning to read and reading to learn are two complex processes for learners with reading comprehension problems. Ying (1995) stresses three possible outcomes from the comprehension process: the reader may construct an interpretation which matches that intended by the author, the reader may construct a satisfactory interpretation of the text which differs from that intended by the author, in which case there may be ambiguity in the surface structure of the text or the text may be ambiguous in the pragmatic sense, and finally, the reader may fail to construct an interpretation of the text. This may be due to its surface structure or to the lack of appropriate knowledge on the part of the reader. Saarnio, Oka, and Paris (1990) believe that decoding skills are thought to be particularly important in comprehension among young children. In addition based on Kintsch and Kintsch (2005) working memory span impacts the understanding of text. When the working memory is limited, the processing of text is constrained, thus impacting understanding. Other factors affecting reading comprehension include knowledge of words and general knowledge (Smith, 1994).

Validity of Reading Comprehension Tests
The notion of test validity has been employed in the testing field for a long time. As a result, the concept of validity has been reviewed many times and has various definitions. According to traditional definition of validity given by Kelly (1927) the problem of validity is that of whether a test really measures what it supposes to measure. This definition seems a little too general and vague. Weir (2005) proposes a narrower definition of validity as the extent to which test scores reflect test takers’ true level of language knowledge or skills. He suggests that validity is found in the scores of a given test instead of the test itself. He also claims that validity is a multifaceted concept and that various types of evidence are needed to establish overall validity of a given test. Some of the major types of test validity are; content validity, whether the test
measure an ample portion of the subject matter, construct validity, whether the test measures the traits it is designed to measure, predictive validity, assesses test takers' likelihood of future success, consequential validity, encompasses various consequences of a test, and face validity, whether the test takers perceive the test as relevant and a good measure of their ability. The concept of construct validity has become so important in language testing that it has often been used as a superior term to all other kinds of validity. Bachman (1990) states that:

“construct validity concerns the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs” (p.225).

The Nature of Synonyms and Opposites

Synonyms are two or more forms with very closely related meanings, which are often, but not always, inter-substitutable in sentences (Yule, 1996). He notes that the idea of sameness of meaning is not necessarily total sameness. According to Fromkin, Rodman, and Hyams (2003) there are no perfect synonyms, i.e., no two words ever have exactly the same meaning. Yule (1996) defines antonyms as two forms with opposite meanings. They are divided into two main types; gradable and non-gradable antonyms. Gradable antonyms can be used in comparative constructions, such as the pair, big/small. Non-gradable antonyms are also called complementary pairs such as the pair, dead/alive. Fromkin and et al (2003) refer to auto-antonyms as the words that are their own antonyms. According to Robinson (1976) a synonym is a word or brief expression close in meaning to the unknown or hazy language unit. According to Blachowicz, Fisher, and Watts-Taffe (2005) it is important for students to learn the difference between the denotations and connotations of words. The denotation of word is its general or literal meaning, and the connotation of a word is what may be suggested by or associated with the use of the word. Powell (1986) argues that the use of antonyms can be one of the most powerful tools in vocabulary instruction. He notes that semanticists identify three main types of opposition; contradictories, complementaries, contraries, and reciprocal, converse, terms. Contradictories are mutually exclusive, such as the pair single/married. Contraries allow for gradations, such as the pair big/small. In reciprocal terms, one word reverses or undoes the meaning of the other, such as the pair buy/sell. However, for instructional purposes, he suggests drawing a distinction between polar antonyms and scalar antonyms. Polar antonyms are categorical and allow no intermediate terms, in other words the assertion of one denies the possibility of the other, such as the pair husband/wife. Both contradictories and reciprocals fall into this category. Scalar terms, in contrast, allow gradations between the extremes, such as big, large, small, tiny. One of the instructional techniques that are possible with scalar terms is a semantic gradient. In this technique, students arrange words on a gradient from one end of the scale to the other. Powell (1986) suggests that when teaching meaning of words, have students use the context in which a word is written to try to work out its meaning, rather than using a dictionary. In other words, have students think about the possible meanings of unknown words by the way words are used in context. This can also be applied to the teaching of synonyms. Rather than approaching a dictionary students should try and work out a word of similar meaning by the context that it is used.

The need for teaching students methods to improve their reading comprehension seems clear. A great deal of students has difficulty with reading comprehension. They look for visual cues within a text rather than an inferred or evaluative response in order to clarify what they have
read. It is vitally important for students to have some level of reading comprehension. According to Pressley (2000) several studies have demonstrated that students' understanding and recollection of the text can be enhanced through explicit teaching in comprehension skills. Miller and Calfee (2004) believe that students' lack of understanding will impinge on their ability to process their reading and a limited vocabulary does not enable them to successfully respond to what they have read.

The purpose of this study is to investigate the effect of teaching within-text key word synonyms, opposites and related words on performing on reading comprehension of TOEFL, among Iranian EFL learners.

METHOD
The participants of the study were 60, 23 males and 37 females, learners of English as a foreign language in the city of Mashhad, Iran. All the participants were advanced learners of English who attended a TOEFL preparation course.

In the current research the following tests were employed:
- To homogenize the subjects of the study and in order to ensure that the members of two groups belonged to the same population, a proficiency test which was a truncated version of the Test of English as a Foreign Language (TOEFL, published by ETS, 2010) was used. The test consisted of 40 structure and written expression items ad 50 reading comprehension items.
- A reading comprehension test was used as the pre-test. The test was extracted from TOEFL iBT sample tests by the researcher. The average difficulty level of the reading passages of the post-test with the utilization of the Fog Index Formula of readability was 16.71.

The procedure will be reported in three consecutive stages to better demonstrate the steps taken in testing the stated hypothesis. In order to homogenize the participants upon their level of proficiency and their reading ability, first a TOEFL (described in full earlier) was administered to all the 60 available students who took part in TOEFL preparation course. They were then randomly divided into experimental and comparison groups. The comparison group consisted of 30 students, 10 males and 20 females, and the experimental group consisted of 30 students, 13 males and 17 females.

As it was mentioned earlier, the experimental and comparison groups, were selected from among a population of students whose proficiency level was examined to be at the same. However, since the study concentrated on reading comprehension ability, a reading comprehension test, adapted from TOEFL iBT sample tests, was administered to both groups to ensure that they were also homogenous in terms of reading comprehension ability.

The participants in the comparison group were taught through a step-by-step procedure which basically focused on teaching reading comprehension strategies to the participants. The strategies taught including; understanding words in contexts, choosing and finding within-text key word synonyms and antonyms, selecting a topic sentence, understanding details and referents, identifying paragraphs and sentences, making inferences, drawing conclusions, and identifying restatements. After these strategies were presented by the instructor, the students practiced them in groups and individually as a class activity and as homework assignment.
The participants of the experimental group were also taught through a process-oriented approach to teaching reading comprehension. The same strategies, described in full earlier, were taught to this group. In addition to other skills the instructor highlighted the importance of learning and practicing within-text key word synonyms and opposites. Before each session the instructor made a list of within-text key word synonyms, opposites and their meanings, the lists were available to all students of the experimental group. Each session the instructor taught the list of words and asked the students to find and circle the key words which were synonyms and had the same or similar meaning, within the text. The repeated words were not considered synonyms. The instructor did the same with within-text key word opposites. Different learning activities such as matching within-text key word synonyms and opposites, filling in the blanks, were used as classroom activities. In order to make sure that the students have learned within-text key word synonyms and opposites, the instructor selected some TOEFL iBT reading passages, including within-text key word synonyms and opposites, as homework assignment.

After ten weeks of teaching reading comprehension through a process-oriented approach to the both comparison and experimental groups, respectively, a reading comprehension post-test, adapted from TOEFL iBT sample tests, was administered to both groups in their final examination, so that the researcher could verify the stated hypothesis of the study through the analysis of the obtained data.

**FINDINGS AND DISCUSSION**

After administrating the TOEFL to 60 students, descriptive statistics was obtained. They were then randomly divided into experimental and comparison groups.

Since the study concentrated on reading comprehension ability, and to check the homogeneity of the experimental and comparison groups at the beginning of the experimentation, a reading comprehension test was administered to both groups as a pre-test, and an independent t-test was performed on the two groups test scores. The mean scores for the experimental and comparison groups on the pre-test were 34.57 and 34.17 respectively which resulted to a difference of 0.40 score.

Regarding the mean scores of two groups there was no significance difference, but in order to be sure of close homogeneity of two groups, a t-test was run. As it has been shown in table 1, the t-observed value for the comparison of the means of two groups was 0.129 at 58 degrees of freedom, which was lower than the t-critical of 2.000. Thus it could be claimed that the two groups were not significantly different in terms of reading comprehension before undergoing the treatment.

Table 1. Comparison between variances and means of the two groups on the reading pre-test

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>T</td>
</tr>
<tr>
<td>Pre-test</td>
<td>Equal variances assumed</td>
<td>11.596</td>
<td>.001</td>
</tr>
</tbody>
</table>
Table 1. Comparison between variances and means of the two groups on the reading pre-test

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>T</td>
</tr>
<tr>
<td>Pre-test</td>
<td>11.596</td>
<td>.001</td>
<td>-</td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>-</td>
<td>.129</td>
<td>49.773</td>
</tr>
</tbody>
</table>

After ten weeks of instruction, both groups were given a similar post-test. The mean score of experimental group was 44.67 which was greater than the mean score of comparison group, 36.73.

To see whether the treatment was effective or not, the means of two groups were compared through a t-test. As it has been shown in Table 2, the t-observed value was 4.263 at 58 degrees of freedom which was higher than the t-critical of 2.000. Thus, the null hypothesis could be safely rejected at 0.05 level of significance, and it could be claimed that the treatment was effective enough to make a significant difference between the experimental and comparison groups.

To assure that the difference between the means of two groups in post-test is related to the instruction of within-text key word synonyms and opposites, the researcher analyzed the questions of both pre and post tests in Table 3.

In order to compute the significance of the differences among the means of the three groups of the questions in both pre and post tests, a One-way ANOVA was performed.

Table 4 represents that the sigma value in experimental group in post-test is 0.00 which is lower than the critical value of 0.05, and the Tukey B. (Table 5) shows the significance of contrasts amongst the groups the means of synonyms and opposites questions (0.96) is significantly higher than the means of other groups (0.74 and 0.76). The experimental group showed significant difference only with the synonyms and opposites questions and surprisingly the comparison group did not show any significant difference with the questions of both pre and post tests. On the whole, as the means of gained scores indicated the experimental group only outperformed the comparison group in synonyms and opposites questions. Consequently, it could be concluded that teaching within-text key word synonyms, opposites and related words would certainly improve students’ reading comprehension ability.

Since the study concentrated on reading comprehension ability, a reading comprehension test was administered to both groups as a pre-test. According to this test the mean score of the experimental group was 34.57 and that of the comparison group was 34.17. Based on Best and Kahn (2006), regarding the mean scores of the two groups, there was no significant difference between the groups and the distribution of samples was normal (a difference of 0.40 score between the man scores is not significant, and the p-value of two groups was higher than the index of 0.05). In order to be sure of close homogeneity of the groups, a t-test was run. The t-observed value for the comparison of the means of two groups was 0.129 at 58 degrees of freedom, which
The effect of teaching within-text key words on students’ performance

was lower than the \( t \)-critical of 2.00 (\( t \geq 2.00 \)) as mentioned in Best and Kahn (2006). Thus it could be claimed that the two groups were not significantly different in terms of reading comprehension ability before undergoing the intervention.

Table 2. Comparison between variances and means of the two groups on the reading post-test

<table>
<thead>
<tr>
<th></th>
<th>Levene’s Test for Equality of Variances</th>
<th>( t )-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F )</td>
<td>Sig.</td>
<td>( T )</td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>16.194</td>
<td>.000</td>
<td>-4.263</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>-4.263</td>
<td>44.051</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 3. Pre and post tests question types

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Test</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonyms &amp; Opposites</td>
<td>18</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Inferencing</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>18</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. One-way ANOVA of the post-test of the experimental group

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>( F )</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>.866</td>
<td>2</td>
<td>.433</td>
<td>8.930</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>4.219</td>
<td>87</td>
<td>.048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5.085</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Multiple comparison of the post-test questions of the experimental group

<table>
<thead>
<tr>
<th>Question Type</th>
<th>N</th>
<th>Subset for alpha = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Others</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Inferencing</td>
<td>30</td>
<td>.7433</td>
</tr>
<tr>
<td>Synonyms &amp; opposites</td>
<td>30</td>
<td>.7617</td>
</tr>
</tbody>
</table>

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 30.000.

After the instruction, both groups were given a similar post-test. As it was mentioned earlier, the mean score of experimental group, 44.67, was greater than the mean score of the comparison group, 36.73. Regarding the mean scores of two groups, there was a significance difference between the performances of the participants of both groups, that it could be due to the intervention. To see whether the instruction was effective or not, the means of two groups were compared through a \( t \)-test. The results of \( t \)-test showed that the \( t \)-observed value was 4.263 at 58 degrees of freedom which was higher than the \( t \)-critical of 2.00 (\( t \geq 2.00 \)). Thus, it could be claimed that the instruction was effective.
and the difference between the groups was due to the effectiveness of instruction.

To determine the effect of teaching within text key word synonyms, opposites, and related words, a matched t-test was also conducted for the mean achievement of both experimental and comparison groups. Regarding the mean scores of the participants of the comparison group in both pre and post tests, the t-observed value was 1.780 at 29 degrees of freedom which was lower than the t-critical value of 2.045 at 0.05 level of significance (t≤0.045). Thus, it could be claimed that there was no statistically significant difference between the pre and post tests mean scores of the participants of the comparison group, who were not exposed to within-text key word synonyms, opposites, and related words.

Regarding the mean scores of the participants of the experimental group in both pre and post tests, the t-observed value was 3.924 at 29 degrees of freedom which exceeded the critical value of 2.045 at 0.05 level of significance (t≥2.045). Based on these findings it was concluded that the performance of the participants of the experimental group in post-test was better than that of the pre-test, supporting the argument that whether the learners are exposed to within-text key word synonyms, opposites and related words, their reading comprehension ability will highly improve.

To assure that the difference between the means of two groups in post-test is related to the instruction of within-text key word synonyms, opposites, and related words, the researcher divided the questions of both pre and post tests into three groups of synonyms and opposites questions, inferential questions, and other types of questions. Then the researcher computed the significant differences among the means of these groups of questions in both pre and post tests by performing one-way ANOVA. As the final step and due to the significant of differences among the means, the post-hoc Tukey B test was calculated. The value of sigma in pre-test questions of the comparison group was 0.074 which was higher than the critical value of 0.05 (sig≥0.05), and the mean ranges of three groups, 0.64, 0.66, and 0.75, at the level of 0.05 were within the same group. It boils down to the fact that the students’ performances, in comparison group, on three groups of questions in pre-test were statistically at the same level. The value of sigma in post-test questions of the comparison group was 0.195 which was higher than the critical value of 0.05 (sig≥0.05), and the mean ranges of three groups, 0.68, 0.71, and 0.78, at the level of 0.05 were within the same group. Thus, comparing the findings of the comparison group on both pre and post tests questions, the researcher concluded that there was no significant difference between the means of three groups of questions and students’ performances on them were statistically at the same level. The value of sigma in pre-test questions of the experimental group was 0.06 which was higher than the critical value of 0.05 (sig≥0.05), and the mean ranges of three groups, 0.57, 0.75, and 0.75, at the level of 0.05 were within the same group. It was concluded that the students’ performances in experimental group, on three groups of questions in pre-test were statistically at the same level. The value of sigma in post-test questions of the experimental group was 0.00 which was lower than the critical value of 0.05 (sig≤0.05), and the mean ranges of three groups, 0.74, 0.76, and 0.96, at the level of 0.05 were not within the same group. The mean range of 0.96, which was related to synonyms and opposites questions, was significantly higher than the means of other groups. The experimental group showed significant difference only with the synonyms and opposites questions and surprisingly the comparison group did not show any significant difference with the questions of both pre and post tests. On the whole, as the means of gained scores indicated, the experimental group only
outperformed the comparison group in synonyms and opposites questions. Consequently, it could be concluded that teaching within-text key word synonyms, opposites, and related words would certainly improve students’ reading comprehension ability.

CONCLUSION
The results of the present study should be interpreted with some caution, due to some methodological limitations. The first limitation of the present study is related to the bias on the part of the experimental group. Since the researcher was the instructor of experimental group he might behave in such a way to affect students’ performance. Another limitation of this study refers to the events which may occur in the learning environment at the same time. These events may affect students’ knowledge of content and skills. And finally the matter of generalizability is considered as the last limitation. Since the researcher had access to a limited number of participants, the generalizability of the findings may decrease.

In every text there are a number of syntactic features called cohesive devices which show, to a large extent, which words and sentences, are related to each other. Halliday and Hassan (1976) stated that these cohesive devices in English include reference, substitution, ellipsis, conjunctives, demonstratives, synonyms, antonyms, hyponyms, metonyms, and repetition. Among these cohesive devices within-text key word synonyms and opposites are two devices which were evaluated and discussed in this study. To sum up, this study was an attempt to explore the potential effect of teaching within-text key word synonyms, opposites and related words on students’ performance on reading comprehension tests, especially TOEFL among Iranian EFL learners. In the present study, teaching of within-text key word synonyms and opposites was conducted in one class. It was meant to see whether performing this instruction in the classroom could improve students’ performances on reading comprehension tests and thus provide students with a new strategy of test taking. As the finding showed, the students in experimental group outperformed those in comparison group which is indicative of the effectiveness of this form of instruction (teaching of within-text key word synonyms and opposites).

The results of the present study will be analyzed along a main dimension; construct validity and reading comprehension tests. Spolsky (1985) suggests that language testing has evolved through three main phases: a first pre-scientific phase utilizing examinations subjectively marked by a single examiner, a second phase focusing on objectivity and reliability, and a third phase which emphasizes validity along with the aspects in the second phase. Unfortunately, in the case of reading comprehension tests, the evolution seems to have largely stopped at the second phase, and when the concept of validity has been addressed, it has typically not been construct validity, but other types such as criterion validity (Henning, 1991).

This study is an attempt to apply this line of reasoning about construct validity to the reading comprehension items to the TOEFL. The purpose of this study is to investigate the effects of teaching within-text key word synonyms, opposites and related words on performing on reading comprehension of TOEFL, among Iranian EFL learners. Subjects are given a number of key word synonyms and opposites and their translated forms of TOEFL items and then tested to discover whether within-text key word synonyms and opposites actually affect their performance on reading comprehension of TOEFL. Based on the findings of the present study, the experimental group outperformed the comparison group in the within-text key word synonyms and opposites questions. Thus, the null hypothesis, there is no
statistically significant difference between the mean score of the students who are exposed to within text key word synonyms, opposites and related words and those who are not, on reading comprehension of TOEFL, among Iranian EFL learners, can be safely rejected. In a similar study, the relationship between TOEFL vocabulary items and meaning, association, collection, and word class knowledge, Schmitt (1999) concluded that TOEFL vocabulary items were able to give only a limited amount of information about the wider range of word knowledge necessary to master a word. The items were not particularly strong in indicating the subjects’ association, word-class and collection knowledge of the target words. Based on Fisk and et al. (2003) teaching synonyms has certainly improved students vocabulary and their meaning making skills in using words in context to work out what they mean. Regarding the nature of teaching synonyms and opposites, Takatsuka (1998) concluded that a different and perhaps more promising approach might be to provide an exhaustive list of possible paraphrase patterns. The upshot is that this study suggests that we need to have a closer look at what items like this are really measuring. This implies that future item format validation should include a prominent construct validity element.

Several steps can be taken in the future to assess and improve the reading validity of the reading comprehension tests. First of all, validity is a multifaceted concept. Therefore, various types of validity evidence can be used in reading comprehension tests in addition to the one explored in the present study. These include, face validity, content validity, concurrent validity, predictive validity. Moreover, in order to receive a more complete construct-related validity evidence of the reading comprehension tests, it is absolutely essential to determine whether the passages used in the test have appropriate levels of text difficulty. Therefore, conducting a text difficulty analysis for reading comprehension tests will result in a more comprehensive evidence of its overall validity.

Second, this study does not address the question of how to ensure that each reading objective receives fair representation in reading comprehension tests. Is there a specific number or percentage of test items each objective should cover? What are some of the possible ways to determine whether an objective is well represented in a test? Finally, answers to these questions and applying them to the validation of the reading tests could significantly strengthen the results of reading tests validations.

Finally, there is another way to improve the future reading tests validation efforts. Conducting test item analysis to determine each test’s item difficulty and discrimination values is another important way to verify the overall validity of the test. It was beyond the scope of the current study to conduct such evaluation. However, conducting this type of evaluation will help the test developers know which reading items are too easy or too difficult for the learners and which items do not discriminate well between learners with higher and lower reading proficiency. Having this information will help test developers improve reading tests validity and overall quality.

Acknowledgement
I would like to express my appreciation to Dr. Babak Shaham for his insightful guidance and support. I also owe thanks to Dr. Jahanbaksh Nikoopour who taught me how to be a researcher.

REFERENCES
The effect of teaching within-text key words on students’ performance


