

Investigating the Relatedness of Cloze-Elide Test, Multiple-Choice Cloze Test, and C-test as Measures of Reading Comprehension

Simin Sattarpour

Ph.D. student, Tabriz University

Parviz Ajideh

Professor, Tabriz University

Abstract

Reading comprehension ability consists of multiple cognitive processes, and cloze tests have long been claimed to measure this ability as a whole. However, since the introduction of cloze test, different varieties of it have been proposed by the testers. Thus, the present study was an attempt to examine the relatedness of Cloze-Elide test, Multiple-choice (MC) cloze test, and C-test as three different types of cloze procedure used for measuring reading comprehension. To this end, one C-test consisting of four short texts, one fixed ratio (n=7) multiple-choice cloze test, and one cloze-elide test were prepared from reading passages with similar readability levels. The participants of the study were 30 (male & female) freshman university students majoring in English literature. The results of ANOVA test showed that there were not any statistically significant differences at the 0.05 level of significance among the performance of the students on the three tests measuring their reading comprehension. Therefore, it was concluded that against the advocates of each test who claim superiority of it over the other types, these three types of cloze tests in this study assessed the reading comprehension in a similar way. So, the testers can be confident to make use of these tests as reading comprehension tests interchangeably.

Keywords: Cloze-Elide test, Multiple-choice cloze, C-test, reading comprehension

تاریخ وصول: ۱۳۹۳/۲/۱ تأیید نهایی: ۱۳۹۳/۶/۸

E-mail: s.sattarpour@yahoo.com

E-mail: parvizaj@gmail.com

Introduction

Reading is a very common and essential activity in everyday life. At the same time, it is a very complex and dynamic process that the reader needs to understand and connect ideas in a text and also use many mental processes to analyze the information. Veeravagu, Muthusamy, Marimuthu, and Subrayan (2010) define reading comprehension as a thinking process by which a reader selects facts, information, or ideas from printed materials; determines the meanings the author intended to transmit; decides how they relate to previous knowledge; and judges their appropriateness and worth for meeting the learner's own objectives (p. 206).

To be able to comprehend what is read, it is needed to be familiar with text structure and topic, to be aware of reading strategies, and to know how to use these strategies in the processing of material and word recognition (Pang, 2008). Reading processes are divided into lower-level and higher-level processes (Grabe & Stoller, 2002). Also, Sadeghi (2007) has identified two factors, internal and external factors, in reading comprehension. While internal factors are related to the reader such as cognitive abilities and strategies, background knowledge, and characteristics, external factors are related to text modality, text characteristics, time and place of reading.

Recently, most researchers (e.g., Carr & Levy, 1990; Grabe, 2009; Koda, 2005, 2007; Stanovich, 2000) emphasize on the component-skills approach to reading. In this approach, reading comprehension consists of multiple cognitive processes (e.g., decoding, vocabulary knowledge, syntactic processing, metacognition) and involves separate measurements of sub-skills of reading. This kind of approach lets us observe how different processes interact with one another such as Stanovich's (2000) interactive-compensatory model of reading, and also it shows how much individual and collective contribution each process makes to reading comprehension (Jeon & Yamashita, 2014). They also believe that the component-skills approach is useful to language teachers and testers by helping them identify areas of individual ability differences and design effective intervention programs and tests.

Literature Review

Testing Reading Skill

Like language learners who have a difficult job to comprehend a text, language teachers and testers encounter a hard task in assessing this ability and skill of learners. Reading is a receptive skill and the processes that the readers go through which, are not observable to the language testers. Lee (2004) defines reading comprehension test as a series of related items that are based on the same reading passage. However, there are some factors affecting reading comprehension test results. Alderson (2000) argues that selected text and test methods are so effective in testing reading comprehension. As most of the studies on reading tests (e.g. Phakiti 2003; Atai & Soleimany, 2009) have shown, the choice of text has a marked effect on the test scores. Therefore, the language testers need to know how to choose an appropriate text. Hughes (2003) argues that successful choice of texts depends on experience, practice and a certain amount of common sense.

Day (1994) discusses seven factors which should be considered in the selection of texts for reading, but in this study, only one of them i.e. readability is considered in the selection of the text for testing reading. Alderson (2000) points out that the difficulty level of the text is one of the important issues to be considered in the selection of text. If a reader-text mismatch, the result will be the user's frustration and failing to use or ignoring the text (Zamanian & Heidary, 2012). To avoid such mismatch, educators would like a tool to check if a given text would be readable by its intended audience. To this end, readability formulas were originally created to predict the reading difficulty associated with the text. All in all, readability is concerned with ensuring that a given piece of writing reaches and affects its audience in the way that the author intends (Zamanian & Heidary, 2012).

Beside the choice of reading text, test techniques used by testers have an important effect on test results. Brown (2004) has classified reading test techniques as reading aloud, written response, multiple-choice, picture-cued items, matching test, editing, gap filling test, cloze test, C-test, cloze-elide test, short-answer test, ordering test, and

summarizing test. In line with Brown, Heaton (1991) has categorized some testing techniques namely: word matching, picture and sentence matching, true/false reading test, multiple-choice items, completion test, rearrangement items, cloze procedure, and open-ended and miscellaneous items for reading test, while Alderson (2000) has classified them as Multiple-Choice, Cloze Test, Gap-Filling Test, Matching, ordering, Editing, cloze-elide, short-answer, free recall, summary, gapped summary, information-transfer. The techniques used in the current study are multiple-choice cloze test, C-test, and cloze-elide test, and it has been tried to examine the differential performance of the learners on these types of tests.

Cloze-test as a Basic Testing Techniques for Reading Skill

A cloze test can be simply defined as a text of appropriate length and difficulty with every nth word deleted and the students are required to complete by filling in the correct words or their equivalents. Farhady (1986, p. 30) clarifies the cloze procedure as the one in which the test takers are required to fill in the blanks with appropriate words on the basis of contextual clues provided in the passage, and he also contends that cloze tests have probably been the most popular kind of tests. Likewise, Harmer (2002) states that cloze - in its purest form- is the deletion of every nth word in a text (somewhere between every fifth or tenth word). This testing technique is mostly considered appropriate to test reading comprehension. The most common purpose of the cloze test is to measure reading comprehension because it has long been argued that cloze measures textual knowledge: i.e. an awareness of cohesion.

Cloze testing was first introduced by Taylor (1953), who developed it as a reading test for native speakers. The term cloze was derived from the closure concept in Gestalt psychology which means that an individual will be able to complete a task only after its pattern has been discerned. A cloze unit may be defined as any single occurrence of a successful attempt to reproduce accurately a part deleted from a 'message' (any language product), by deciding from the context that remains, what the missing part should be (p. 416). Therefore, Mc Kamey (2006) claims that language skill is not the only skill required to 'cloze' the gaps created by deleted words, but rather a

kind of non-verbal reasoning skill, known in Gestalt psychology as 'closure', is needed. This procedure is the indicator of a tendency of humans to complete a familiar but not-quite finished pattern (Lu, 2006).

Spolsky's reduced redundancy is the major concept which has been exploited in cloze procedure (Spolsky, 1969). Language reduced redundancy (LRR) approach to language testing was firstly proposed by Spolsky and has been utilized in developing different test procedures, such as cloze test, C-test, dictation, etc. Spolsky believes that the knowledge of language requires the ability to act properly even when there is reduced redundancy; that is "language learner presented with mutilated language" (p. 79) can use his/her acquired competence to restore either the original text or an acceptable text or restore the message in the noise tests (Klein-Braley, 1985). Hence, Spolsky argues that knowing a language equals the ability to grasp an incomplete or distorted message and make educated guesses about a certain percentage of the missing information. They have also claimed that the more proficient a learner is, the higher scores he can get on a LRR test. According to Klein-Braley (1997), cloze test, is the most important and best-known operationalization of the LRR principle.

Both teachers and researchers continue to use means derived from the distributions of cloze scores to compare the readability of texts, the language proficiency of students, and the intelligibility of a given author (Oller & Jonz, 1994, p. 13). There are, however, at least five main types of cloze tests available to language teachers: The fixed-rate deletion, the selective deletion (also known as the rational cloze), the multiple-choice cloze, the cloze elide and the C-test (Ikeguchi 1995; Klein-Braley & Raatz, 1984). In this paper the last three varieties are touched upon.

1. *Multiple-choice Cloze.* Since the construction of standard cloze tests, some researchers (e.g. Cranney, 1972) have pointed out some demerits of this test technique and they believed that some modifications are required. These researchers argued that the standard cloze is an extremely difficult and anxiety-invoking test, and the required length of a cloze passage makes it inconveniently long as a unit of test assembly. Moreover,

the standard cloze test has the notable disadvantage of requiring hand scoring.

As a modification to standard cloze test, Multiple-choice (MC) cloze test was first suggested by Jonz in 1976. MC cloze tests are a modified form of cloze test in which the deleted words are substituted with a number of possible choices instead of just using a blank like in traditional completion cloze tests. Deleting words can be based on fixed ratio cloze or rational cloze. In fact this type of cloze is a combination of traditional cloze procedure and MC test (Hinofotis & Snow, 1980). The rationale behind the construction of these tests was that whether it was possible to construct a reliable and valid cloze test that could be machine scored and still retained the essential elements of the cloze procedure (Cranney, 1972). Farhady (1996) points out the advantage of MC cloze tests over the standard cloze tests in a way that the task of comprehension is easier than that of production and students will get higher scores than on an open ended form. The excessive difficulty and ambiguity of the original cloze testing situation appears to have been considerably reduced (O'Reilly & Streeter, 1977).

For constructing a MC cloze, like a normal cloze test, an appropriate passage is chosen first. The deletion procedure starts from the second sentence. Every nth (5-10) word is deleted. The second stage requires the supplement of the deletion with three or more distracters. The examinee then chooses from these the word which fits the context best. Alderson (1990) found that providing choices for the deletions lessens the testee's memory load and makes the test taking process easier. O'Reilly and Streeter (1977) mention that the MC cloze test preserves many of the advantages of the cloze technique while enhancing its applicability as a measure of reading comprehension, and also it suffers less from such sources of invalidity as test anxiety because the test passages do not function as ordinary test items until the student reaches the point of no-comprehension with a passage (p. 6).

Hale et al. (1988) have mentioned numerous research studies investigating the correlation of MC cloze test with different standard language proficiency tests (e.g. TOEFL), standard cloze test and vocabulary and grammar tests. These studies have been done on both native and non-native speakers. Based on these studies and their own research, Hale et al. have concluded that an MC cloze test taps processes related to several aspects of English proficiency, including reading comprehension, vocabulary, grammar, and receptive aspects of writing ability.

2. *C-test*. C-test proposed by Klein-Braley in 1980s is an alternative to cloze test and the letter C stands for *Cloze* to call to minds the relationship between the two tests. It usually involves four to six short passages in each of which second half of every second word in each text is deleted and the first and the last sentence of the passage are remained intact which is called rule of two (Klein-Braley&Raatz,1984). Likewise, Khodadadi and Hashemi (2011) claim that the rule of two or principle is the defining feature of C-test. However, in a study, Jafapur (1999) used different deletion rates and deletion starts and showed that there is nothing magical about the rule of two because obtained results was more or less similar. Like all varieties of cloze procedure, the rationale behind the C-test is the reduced redundancy principle.

Although C-Tests are a type of cloze tests, they are superior to them because they enjoy the following six advantages from Klein-Braley (1997) study:

1. Many more items are possible with much shorter texts. A classical cloze test using a 5th word deletion rate would have to be at least 500 words long to contain 100 items. A C-Test consisting of five texts with 20 half-deleted words would be only approximately half as long

2. C-Test scoring is exact and objective because there is almost always one possible solution. In a few cases more than one solution is possible, but almost always only two alternatives. When this happens all possible solutions are counted as correct.

3. The scoring of a C-Test is quick and easy for the native speaker or the teacher since it takes only slightly more time than is needed for simply reading the text.

4. C-Tests are very easy for native speakers. On the other hand, someone who does not understand the language at all normally makes a score of zero or close to zero.

5. Since every second word is damaged the probability of obtaining a representative sample of all the word classes in the text is very much higher.

6. Because the C-Test consists of a number of different texts the sampling of content classes is better (p. 65).

However, as Lee-Ellis (2009) has mentioned in his work, there has been some controversy regarding the C-Test as a measure of proficiency. Proponents of this testing technique praised its high reliability and concurrent validity indices (e.g., Eckes & Grotjahn, 2006; Grotjahn, 2006, 2010; Klein-Braley, 1997; Sigotte, 2004), ease and efficiency of test administration as well as objectivity of scoring (e.g., Klein-Braley & Raatz, 1984), and its alleged measure of integrative use of language (e.g., Dörnyei & Katona, 1992; Klein-Braley, 1997). On the other hand, it also has been subject to criticism for its lack of face validity (e.g. Kontra & Kormos, 2006), poor item discrimination (e.g., Cleary, 1988), and unclear construct validity (e.g. Jafarpur, 1995). (See Babaii & Ansary (2001) for further discussion.)

3. *Cloze-Elide Tests*. Another variety of cloze test is known as the cloze-elide test. This testing technique was proposed by Alderson (2000). The elide technique was introduced by different names such as intrusiveword technique, text retrieval, text interruption, dotted text, mutilated text, and negative cloze (Alderson, 2000, p. 225). This test got its name because originally test takers were expected to cross out the superfluous words on paper, which was called eliding (Manning, 1987).

The cloze-elide (Manning, 1987; Bowen, 1978; Elder & Von Randow, 2008) is an objective language test whereby superfluous, incorrect words are inserted into a text and must

be identified by the test taker within a limited time. Likewise, Brown (2004) defines the cloze-elide test as a kind of test that inserts words to a text that actually do not belong to the text. The test-taker's task is to detect and cross out the intrusive words (p. 204). Manning (1987) compared scores of more than 1,200 ESL students in U.S. universities and found that the elide test was useful test as a reliable and efficient predictor of other English proficiency measures, such as TOEFL scores, graded essay scores, and teacher judgments of student proficiency.

Farhady (1996) has drawn the testers' attention to three points in constructing Cloze-Elide Tests. The first point is the selection of the passage. In selecting the passage, all the rules about the difficulty and the length should be met. The second point is determining the locations where the extra words should be inserted. He suggests the random insertion procedure as the most appropriate method of inserting the words in the text. That is, the words in the passage should be numbered. Then the numbers should be randomly selected and the redundant word should be inserted after or before the word that corresponds to the randomly selected number. It is important to be sure that the inserted words do not belong to the text. Otherwise, the test takers will not be able to identify the inserted words. The third point is the selection of the words to be inserted in the passage. Again, Farhady prefers random selection of words from dictionary.

However, some researchers (e.g. Manning, 1987; Bowen, 1978) have reported some weaknesses with random insertion of words. Firstly, as Bowen (1978) points out, all insertions are not equal: some insertions are very conspicuous, while others manage to partially conceal themselves (p. 3). In addition, Bowen also mentions that with random insertion, by chance some words would be inserted into reasonable places without need for deletion. Inserted words must damage the grammatical or lexical integrity (Bowen, 1978, p. 14) of the sentence so that test takers are not asked to remove some words that are clearly ungrammatical and others that are only

deemed not necessary because they did not happen to be in the original text.

Another issue regarding word insertion is whether items necessitate the examination of adjacent or nonadjacent data to judge appropriateness. Of course, some scholars believe that the inserted word should be similar to the adjacent words. Baker (2011) believes this is a similar concern in traditional rational-deletion cloze. As Hudson (2007) notes about the rational-deletion cloze, the source of item difficulty involves relatively short-range grammatical constraints—usually a few words on either side of the blank or within a single grammatical phrase or clause (p. 102).

Moreover, Scoring techniques of the cloze-elide test is not straightforward. There are some differences in scoring procedures utilized by researchers. In Palupiningsih (2011) study, the test takers were given a credit to the words that were detected and crossed out correctly, but they were penalized for crossing out the words that are not redundant and are parts of the text. On the other hand, Bowen (1978) noted errors of omission and commission in his scoring of this test, which he called insufficient vs. superfluous editing (p. 2). Manning (1987) calculated a score which corrected for guessing and accounted for errors of both omission and commission. Elder and von Randow (2008) and Baker (2011) counted errors of omission only (i.e., they didn't take point off if a test taker deleted a word they didn't need to).

The abovementioned three tests, like cloze procedures, have been mostly used as tests of language proficiency and rarely as tests of reading comprehension. The present study aims to study the effectiveness of the Multiple-Choice cloze tests, C-tests, and the Cloze-Elide tests as measures of reading comprehension. Moreover, as it is mentioned above, there are some controversies among researchers regarding the viability of different varieties of cloze procedures, and the advocates of each specific type of cloze procedure claim that it is better tool for measurement of either language proficiency or reading comprehension. To date very few studies have touched upon this

issue, and also the existing research studies have just focused on two types of cloze procedures which are mostly C-tests , MC cloze, and standard cloze. So in the current study, it is attempted to answer the following research question and test the following null hypothesis:

RQ: Is there any significant difference in the learners performance on the Multiple-choice cloze tests, C-tests, and the Cloze-Elide Tests as measures of reading comprehension?

NH0: There is no significant difference in the learners performance on the Multiple-choice cloze tests, C-tests, and the Cloze-Elide Tests as measures of reading comprehension.

Method

Participants

The participants of the present study were 34 freshmen including 15 males and 19 females with age range of 19-30 majoring in English literature at Tabriz University. However, not all the students were present in all three sessions in which the tests were given and thus the scores of only 30 participants (12 males and 18 females) were used for statistical analysis. Most of them were native Persian or Azeri speakers. Because of some limitations, the researcher was not able to give a placement or proficiency test at the beginning of the study, but to make sure of their homogeneity , the participants scores on the reading comprehension course in previous semester were checked out and the ones with too high or too low scores were excluded from the study.

Instruments

Three varieties of cloze test (Multiple-Choice cloze, C-test, and Cloze-Elide test) were used in this study. The difficulty levels of the tests were in accordance with the proficiency level of the subjects. They were calculated by using Flesch Readability Formula. The excerpts from the texts used in these test are presented in Appendix.

Test Preparation

For preparing the tests, different authentic texts were studied and finally two appropriate ones with the length of about 500-600 words

were selected for MC cloze test and Cloze-Elide test. In addition, four shorter texts were chosen for the C-test. The length of each of these four texts was about 150-170 words. In the selection of texts, it was tried to take into consideration the factors proposed by Day (1994), esp. the readability, the culturally suitable, and appearance factors. The readability of selected texts was calculated by Flesch Readability Formula. The selected difficulty levels for the MC cloze test and Cloze-Elide test were 62.9 and 63.7 respectively. Also, the average of calculated difficulty levels of four texts in C-test was about 63.2. Based on Flesch Readability Formula, the text with these levels of difficulty are called standard or average texts.

Multiple-Choice cloze test: The MC cloze test was developed out of the passage taken from authentic sources (TOFEL) by using a 7th deletion random cloze test. To follow Oller and Jonz (1994), the first and second sentences were left intact, and the deletions began with the 7th word of the second sentence. The difficulty level of the text used for constructing MC cloze test was 62.9. The constructed tests yielded 46 items. The prepared test was piloted twice among 12 learners similar to the sample. At the first time, they were asked to complete the text with appropriate words like standard cloze test. By doing so, the most frequent incorrect responses with the same part of speech written in the pilot were selected as distractors of the MC cloze. In the second time piloting, the malfunctioned distractors were recognized and replaced with suitable ones. Additionally, the required time for test completion was estimated.

C-test: The C-test like cloze test was constructed out of four short passages extracted from authentic sources with approximate difficulty level 63.2. The C-test was constructed by using Klein-Braley (1981) *rule of two*. The first and last sentences of the each text were left intact and the deletion began from the second word of the second sentence. The second half of every second word was deleted and if a word had only one letter it was skipped. Also, if a word had an odd number of letters, the larger half was removed. The constructed C-test had 105 items. The test was also piloted to detect the probable problems and to estimate the length of the time needed to complete the tests.

Cloze-Elide test: Like two other tests in this study, the text for Cloze-Elide test was an authentic text with the level of difficulty of 63.7. It was tried to choose the superfluous words in the same frequency range as the original passage, and no technical and specialist words were inserted. Furthermore, the proportions of function and content words were approximately equal due to the results of eye-tracking research that indicates that readers fix content words much more frequently than function words (Just & Carpenter, as cited in Baker, 2011). Because of some imperfections of random insertion of the words mentioned above, superfluous words were not simply inserted randomly. Following Baker (2011), they were initially inserted randomly upon the line, with one superfluous word per line, then adjusted to make sure words were damaging grammatically and not too conspicuous by their placement (like the last word of a paragraph), and to make sure they were varied in their syntactic placement. This test was also piloted before the administration and the required time for completion was estimated. For example, if some items at the end of the test were not attempted by most test takers it could be evidenced that they need more time to complete the test.

Administration and Scoring of the test

The final versions of the tests were administered with one week interval. Therefore data collection lasted three successive weeks. The allocated time for completing the each test was 25-30 minutes. Before starting to do the test, the participants were provided with a complete instruction both in English and in Persian. In order to respect the anonymity right of the participants, they were told that they are free to write whether their own names or any other pseudonym.

The scoring method for MC cloze test and C-test was exact word method. It can be said that in all three tests the scoring method was objective, so that obtained scores are reliable. The MC test was scored like usual multiple-choice items and each item had one point. In the C-test also each item had one point. In the current study it was decided to tolerate the spelling problems in the C-test which did not changed the meaning and part of speeches of the words. If these two occurred, the written word did not gain any score. Regarding the scoring of Cloze-Elide test, it was decided to apply the same scoring procedure

as Elder and von Randow (2008) and Baker (2011). They counted errors of omission only and didn't deduct any point if a test taker deleted a word they didn't need to, unless they deleted more than one per line, which was against the instructions. In other words, the participants were not penalized for guessing because it is believed that it is not appropriate in the academic context. Incomplete items at the end of a passage were also marked incorrect.

Data Analysis

In order to analyze the obtained data statistically, the SPSS (version 20) software was utilized by the researcher. Moreover, to answer the research question of the study and to establish whether there are any differences in the performance of the participants on three types of cloze tests, besides the descriptive statistics, a one-way ANOVA was computed.

Results and Discussion

After scoring, for the ease of comparison all the scores were calculated out of 100. Descriptive statistics has two functions: (1) it gives a precise description of the characteristics of a score distribution, and (2) it forms a basis for further statistical analyses in finding out the similarities and differences between and among sets of scores (Bachman, 2004). Therefore, descriptive statistics for all three types of cloze test are represented in Table 1 below.

Table 1
Descriptive Statistics for the Scores of the Tests

Tests	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Cloze-Elide	30	60.00	21.03	3.84	9.09	93.18
MC cloze	30	64.71	10.85	1.98	36.96	82.61
C-test	30	65.27	20.29	3.70	20.00	93.33
Total	90	63.32	17.95	1.89	9.09	93.33

As it is shown in Table 1, the mean scores for Cloze-Elide test, MC cloze test, and C-test are 60.00, 64.71, and 65.27 respectively. A

quick look at these numbers indicates that the performance of the participants on MC cloze and C-test was very similar to each other while the participants performed slightly different on the Cloze-Elide test. Figure 1 below clearly indicates that the mean scores of the MC cloze test and C-test were very close to each other while the mean score for Cloze-Elide test was slightly lower than the others.

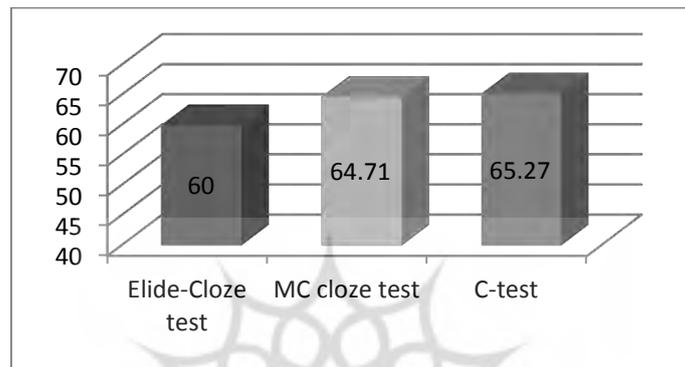


Figure 1. Comparing Means Scores of the tests

However, descriptive statistics alone cannot be used to make inferences from the sample data, so alongside the descriptive statistics, inferential statistics is also needed to make stronger interpretation of the data. In statistics, analysis of variance (ANOVA) provides a statistical test of whether or not the means of several groups are all equal, and therefore generalizes *t*-test to more than two groups, so ANOVAs are useful in comparing two, three or more means.

Before conducting the ANOVA, a normality test was used to know whether or not the distribution of the scores in three tests was normal. Thus, the Kolmogorov-Smirnov test was employed. The data is said to be normal if the *Probability value (p)* is more than 0.05. The result of the normality test showed that the *Probability values (p)* for the Cloze-Elide, MC cloze, and C-test were 0.43, 0.7, and 0.55 respectively. Therefore, it can be claimed that the distribution of all three tests was normal because the calculated (*p*) values are more than .05.

Additionally, the prerequisite for testing the means is the homogeneity of the variances of the scores of the three tests, separately. So, we tested the homogeneity of the variances using the

Levene Test. If the significance level (Sig.) of the test is smaller than 0.05, the homogeneity of the variances will be rejected. As shown in Table 2, the Sig. = 0.150 which is higher than 0.05. Therefore, the homogeneity of the variances is approved.

Table 2
Test of Homogeneity of Variances

Levene Statistics	df1	df2	Sig.
1.274	2	87	.150

After making sure of the normality of distributions and homogeneity of variances, a One-way ANOVA test was utilized to compare the mean scores of the Cloze-Elide test, MC cloze test, C-test. Table 3 below shows the obtained results of this analysis.

Table 3
One-Way ANOVA Results for the Scores of the Tests

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	502.73	2	251.37	.775	.464
Within groups	28202.25	87	324.16		
Total	28704.99	89			

As shown in Table 3, the results indicate that the computed level of significance (0.464) is much higher than 0.05, and thus the differences are not meaningful. Although there were some slight differences among the mean scores of the three tests found in descriptive analysis, the difference between the participants performance on the three types of cloze tests is not statistically significant based on inferential analysis. In other words, the participants performed similarly on the three given tests; thus, the null hypothesis of the present study is confirmed.

As mentioned before, cloze test is the most important and best-known operationalization of the LRR principle. The rationale behind RRP is that natural languages are redundant in the sense that speakers of a language can supply missing linguistic items or they are able to do without them to a certain extent. This characteristic of human

verbal communication system decreases the possibility of making errors and permits communication where there is some interference in the communication channel (Spolsky, 1973). All tests constructed on the basis of RRP exploit authentic materials, damage them in some way, and present them to the examinees to be restored. Thus, the similarities in the results of the three types of cloze tests in this study may be due to the fact that they are different appearances of the same RRP approach.

Conclusion

The current study investigated whether there are any significant differences between the learners performance on the MC cloze tests, C-tests, and the Cloze-Elide tests as measures of reading comprehension. The results of the study indicated that the Cloze-Elide test, MC cloze test, and C-test can be most likely used interchangeably to measure the English reading comprehension skill of the students due to the fact that there were no statistically significant differences between the mean scores of these three tests at the 0.05 level of significance (Sig. > 0.05). Therefore, these results can be taken as obvious evidence that using these three techniques of testing reading comprehension can have the same results; and that they assess the same construct.

The present research can have some implications for both language testers and teachers. Based on the obtained results, they can be confident in using Cloze-Elide test, MC cloze test, and C-test interchangeably as measures of reading comprehension. This study also suffers from some limitations. One of them is that the data of the study was confined merely to 30 subjects. Another limitation to the present study it was conducted on one proficiency level. Therefore, further study should be done on these types of cloze tests as measures of reading comprehension with more subjects and with different levels of language proficiency. Also, it would be very revealing to investigate students attitudes towards the strategies they use while doing these tests by accompanying a questionnaire to the study.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge, Cambridge University Press.
- Atai, M. R., & Soleimany, M. (2009). On the effect of text authenticity & genre on EFL learners' performance in C-text. *Pazhuesh-e Zabanha-ye Khareji*, 49, 109-123.
- Babaii, E., & Ansary, H. (2001). The C-Test: A valid operationalization of reduced redundancy principle? *System*, 29(2), 209° 219.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Baker, B. A. (2011). Use of the cloze-elide task in high-stakes English proficiency testing. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 9, 1-16.
- Bowen, J. D. (1978). The identification of irrelevant lexical distraction: An editing task. *TESL Reporter*, 12(1), 1° 3, 14° 15.
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. White Plains, NY: Pearson Education.
- Carr, T. H., & Levy, B. A. (1990). *Reading and its development: Component skills approaches*. Orlando, FL: Academic Press.
- Cranney, A. G. (1972). The construction of two types of cloze reading tests for college students. *Journal of Literacy Research*, 5(1), 60-64.
- Day, R. R. (1994). *Selecting a passage for the EFL reading class*. *Forum*, 32(1). Retrived from <http://eca.state.gov/forum/vols/vol32/no1/p20.htm>
- Dörnyei, Z., & Katona, L. (1992). Validation of C-Test among Hungarian EFL learners. *Language Testing*, 9, 187° 206.
- Eckes, T., & Grotjahn, T. (2006). A closer look at the construct validity of C-Tests. *Language Testing*, 23, 290° 325.
- Elder, C., & Von Randow, J. (2008). Exploring the utility of a web-based English language screening tool. *Language Assessment Quarterly*, 5(3), 173° 194.
- Farhady, H. (1986). Fundamental concepts in language testing (5) theories. *Roshd Foreign Language teaching Journal*, 2(4), 28-37.

- Farhady, H. (1996). Varieties of cloze procedure in EFL education. *Roshd Foreign Language Teaching Journal*, 12, 217-229.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Grabe, W., & Stoller, F.L. (2002). *Teaching and researching reading*. Harlow, Essex: Pearson Education.
- Grotjahn, R. (2006). *Der C-Test: Theorie, Empirie, Anwendungen/The C-test: theory, empirical research, applications*. Frankfurt/ M: Peter Lang.
- Grotjahn, R. (2010). *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from Current Research*. Frankfurt am Main: Peter Lang.
- Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M., Butler, F. A., & Oller, J. W. Jr. (1988). *Multiple-choice cloze items and the test of English as a foreign language*. Princeton, New Jersey: Educational Testing Service.
- Harmer, J. (2002). *The Practice of English Language Teaching* (3rd Edition). England: Longman
- Heaton, J. B. (1991). *Writing English Language Tests*. New York: Longman.
- Hinofotis, F. B., Snow, B. G. (1980). An alternative cloze testing procedure: Multiple-choice format. In J. W. Oller, Jr. & K. Perkins (Eds.), *Research in language testing* (pp. 129-133). Rowley, MA: Newbury House.
- Hudson, T. (2007). *Teaching second language reading*. New York: Oxford University Press.
- Hughes, A. (2003). *Testing for language Teachers* (2nd Ed.). UK: Cambridge University Press.
- Ikeguchi, C. (1995). Cloze testing options for the classroom. In J. D. Brown, & S. Yamashita (Eds.), *Language testing in Japan* (pp. 166-178). Tokyo: The Japan Association for Language Teaching.
- Jafarpur, A. (1995). Is C-Test superior to cloze? *Language Testing*, 12, 194° 216.
- Jafarpur, A. (1999). Can the C-test be improved with classical item analysis? *System*, 27, 79° 89.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160-212.

- Jonz, J. (1976). Improving on the basic egg: The M-C cloze. *Language Learning*, 26, 255-256.
- Khodadady, E., & Hashemi, M. (2011). Validity and C-test: The role of text authenticity. *Iranian Journal of Language testing*, 1(1), 30-41.
- Klein-Braley, C. (1981). Empirical investigations of cloze tests. PhD thesis. University of Duisberg.
- Klein-Braley, C. (1985). A cloze up on the C-test: A study on the constructive validation of authentic tests. *Language Testing*, 2, 76-104.
- Klein-Braley, C. (1997). C-test in the context of reduced redundancy testing: an appraisal. *Language Testing*, 14, 47-84.
- Klein-Braley, C., & Rattz, U. (1984). A Survey of research on the C-test. *Language Testing*, 1(2), 134-146.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. New York: Cambridge University Press.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57, 1° 44.
- Kontra, E. H., & Kormos, J. (2006). Strategy use and the construct of C-Test. In R. Grotjahn (Ed), *Der C-Test: Theorie, Emprie, Anwendungen/The C-Test: Theory, empirical research ,applications*(pp. 121-138). Frankfurt/am: Peter Lang.
- Lee, Y. (2004). Examining passage-related local items dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing*, 21(1), 74-100.
- Lee-Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing*, 26 (2), 245° 274.
- Lu, G. (2006). Cloze tests and reading strategies in English language teaching in China.(Unpublished MA Thesis). University of Western Cape, China.
- Manning, W. H. (1987). *Development of cloze-elide tests of English as a Second Language*. (TOEFL Research Report 23). Princeton, NJ: Educational Testing Service.
- McMamey, T. (2006). Getting closer on the cloze: A validity study of the 'rational deletion' method. *Second Language Studies*, 24(2), 114-164.

- O'Reilly, R. P. & Streeter, R. E. (1977). Report on development and validation of a system for measuring literal comprehension in a multiple-choice cloze format: preliminary factor analytic results. *Journal of Reading Behavior*, 9, 45-69.
- Oller, J. W., Jr., & Jonz, J. (1994). *Cloze and coherence*. Cranbury, New Jersey: Bucknell University Press.
- Pang, J. (2008). Research on good and poor reader characteristics: Implications for L2 reading research in China. *Reading in a Foreign Language*, 20(1), 199-217.
- Palupiningsih, A. (2011). Testing reading comprehension using cloze tests and cloze-elide tests among the year-10th students of sman 1 depok in the academic year (Unpublished doctoral dissertation). State University of Yogyakarta, Malaysia.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56.
- Sadeghi, K. (2007). The key for successful reader-writer interaction: Factors affecting reading comprehension in L2 Revisited. *Asian EFL Journal*, 9(3), 198-220.
- Sigott, G. (2004). *Towards identifying the C-Test construct*. Frankfurt/M: Peter Lang.
- Spolsky, B. (1969). Reduced redundancy as a language testing tool. Presented at Second International Congress of Applied Linguistics, Cambridge, England, September, 8-12.
- Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- Veeravagu, J., Muthusamy, C., Marimuthu, R., & Subrayan, A. (2010). Using Bloom's Taxonomy to gauge students reading comprehension performance. *Canadian Social Science*, 6(3), 205-212.
- Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies*, 2(1), 43-53.

Appendix: The instruments of the study

The C-test

In each of the texts below, the second part of every second word has been taken out. The number of the deleted letters equals the number of remaining letters or one letter more. Please fill in the missing parts to complete each text.

Text 1

Fire, water, wind and oil are important sources of energy. They pro - -
- - electricity, co - - food, a - - run t - - machines wh - - - manufacture
wh - - people ne - - . Because t - - world's popul - - - - is incre - - - -
- , and bec - - - - industry a - - - technology a - - growing, addit - - - -
supplies of energy sou - - - - , such as oil a - - gas, a - - hard to get a -
- expensive, how - - - - . Others a - - not y - - efficient. Ther - - - - ,
the devel - - - - of alte - - - - sources of energy is an impo - - - -
goal of today's scien - - - - and techno - - - - . This reading tells
about three alternate sources: solar energy, geothermal energy, and
coal.

Solar energy comes from the sun s heat. La - - - panels a - - used to
collect t - - heat bro - - - - to t - - earth by the ra - - of t - - sun. T - -
heat is then sto - - - in the - - - - mass. It can be us - - to heat wa - - -
and ho - - - and to generate tur - - - - to co - - as a sou - - - of ene - - - .
The Soviet Union, the United States, and China all have large coal
deposits.

Text 2

People in different countries may eat the same food but they prepare it
very differently. For exa_ _ _ _ , Chinese so_ _ is th_ _ and cl_ _ _ , but
Ger_ _ _ soup i_ thick a_ _ heavy. So_ _ people li_ _ raw me_ _ ,
while oth_ _ _ like me_ _ only i_ it is well-cooked. Ma_ _ people li_ _
butter fr_ _ _ and fi_ _ , but th_ _ _ are peo_ _ _ in India who like it
melted into an oil before they eat it. Many people in the East like plain
boiled rice, but in some countries people like theirs made into a sweet
pudding.

The MC cloze test

Read the following text and fill the blanks by choosing the most
appropriate word from the choices given at the end of the text.

Nowadays it is very common for mothers to work outside the home.
Whether a woman should stay at __(1)___ or join the work force is

__(2)___ by many people. Some argue that __(3)___ family, especially the small children, maybe ____ (4) _____. The fact is, however, that many __(5)___ need to work because of economic ____ (6) ____ or want to work to maintain __(7)___ career. I believe that every mother __(8)___ the right to work, and the ____ (9) _____ to work should be one that a __(10)_____ makes on her own. But first __(11)___ should carefully consider the many problems __(12)_____ affect mothers who work.

The major ____ (13)_____ a working mother faces concern her ____ (14) _____. She must either find a reliable ____ (15)_____ who will be loving toward the __(16)_____ or a good day-care centre ____ (17)_____ the children can go. If a ____ (18)_____ gets sick, the mother must make __(19)_____ arrangements for the child to be __(20)_____ for at home or she must __(21)_____ home from work. While at work, __(22)___ mother may worry about her children. ____ (23)___ may wonder if they are safe, __(24)___ they are learning the values she __(25)___ them to have, and her absence __(26)___ hurting them emotionally. She may also ____ (27)_____ not being able to take them __(28)_____ after school activities or participates in ____ (29)_____ activities with them.

- | | | | |
|--------------------|----------------|--------------|--------------|
| 1) A. home | B. dorm | C. house | D. farm |
| 2) A. debated | B. ignored | C. imbedded | D. attested |
| 3) A. the | B. of | C. all | D. in |
| 4) A. hurt | B. resigned | C. neglected | D. entitled |
| 5) A. families | B. humans | C. women | D. widows |
| 6) A. factors | B. reasons | C. issues | D. dispute |
| 7) A. the | B. of | C. a(n) | D. at |
| 8) A. has | B. owns | C. bit | D. had |
| 9) A. right | B. decision | C. reason | D. thought |
| 10) A. family | B. husband | C. children | D. referee |
| 11) A. he | B. kid | C. she | D. mob |
| 12) A. that | B. but | C. also | D. here |
| 13) A. possibility | B. problems | C. reasons | D. outcomes |
| 14) A. husband | B. family | C. children | D. parents |
| 15) A. person | B. aunt | C. caregiver | D. center |
| 16) A. mothers | B. women | C. humans | D. children |
| 17) A. whose | B. where | C. which | D. whom |
| 18) A. woman | B. child | C. spouse | D. peeress |
| 19) A. additive | B. comfortable | C. expensive | D. special |
| 20) A. rested | B. cared | C. whined | D. comforted |
| 21) A. stay | B. probe | C. come | D. clean |

- | | | | |
|---------------|-----------|-------------|------------|
| 22) A. a | B. the | C. as | D. such |
| 23) A. they | B. she | C. him | D. her |
| 24) A. or | B. if | C. unless | D. and |
| 25) A. wants | B. thinks | C. launches | D. teaches |
| 26) A. is | B. causes | C. brings | D. results |
| 27) A. worry | B. regret | C. feel | D. rubout |
| 28) A. off | B. to | C. in | D. for |
| 29) A. school | B. family | C. park | D. movie |

The Cloze-Elide test

There are 44 extra words in this text. Each line has one extra word.

You are going to find, and then circle or underline them.

You may could know that around the world people drink more coffee than tea. And that, although after oil, it is the second biggest traded commodity. You may also know that because the drink comes from beans which are first roasted and then ground. But what else do you know about this realizing popular stimulant?

It is said that the story of coffee left started with an Ethiopian goatherd named Kaldi. One day Kaldi was surprised to see that by his goats were behaving very strangely: instead of grazing quietly as normal they were scarcely jumping around, almost dancing. He also noticed the red cherries from a plant that with the goats were eating. He tried some himself and was surprised by the feeling of public extreme happiness and excitement - he felt assisted like dancing too!

But it wasn't used this as a drink at first, but as a food. The coffee berries, mixed having with animal fat, were used by monks to stay awake during long hours of prayer. From Ethiopia coffee was later into cultivated in Yemen and the first hot drink was developed there around on AD 1000. Three centuries later Muslims were keen coffee drinkers be and as Islam spread, so did coffee. Coffee houses appeared material in Cairo and Mecca.